The Microbial Community Ecology of the Cystic Fibrosis Lung

Cuthbertson, Leah Forbes

Awarding institution:
King's College London

The Microbial Community Ecology of the Cystic Fibrosis Lung

A thesis submitted for the degree of Doctor of Philosophy in the Institute of Pharmaceutical Science, Molecular Microbiology Research Laboratory, King's College London

By

Leah Cuthbertson

Pharmaceutical Sciences Research Division

Kings College London

October 2014

Declaration

“I declare that I have personally prepared this report and that it has not in whole or in part been submitted for any degree or qualification. The work described here is my own, carried out personally unless otherwise stated and written in my own words. All sources of information, including quotations, are acknowledged by means of reference.”

Signature:

Date:

Abstract

Respiratory failure, due to infection and concomitant inflammation is the major cause of morbidity and mortality in people suffering from the genetic disorder, cystic fibrosis (CF). Consequently, the CF Foundation currently estimates that patients with CF have a median predicted life expectancy of only 41.1 years. Understanding the relationship between the complex and diverse bacterial community present within the lower respiratory tract and patient outcomes has therefore become a top priority. Through the use of next generation sequencing technologies (Roche 454 and Illumina MiSeq) and ecological statistics and modelling, the complex relationships between the bacterial community within the CF lung and host related clinical factors were investigated.

By first establishing guideline methodologies for the reduction of bias in the collection, storage and treatment of respiratory samples, this thesis aimed to use large scale spatial and longitudinal studies to investigate key relationships between the bacterial community and clinical factors.

It has been well established that a complex and diverse bacterial community exists within the CF lung. Spatial sampling revealed key relationships between the bacterial community and other diagnostic parameters including, FEV_1, gender, and clinic location.

Longitudinal sampling aimed principally to investigate CF pulmonary exacerbations (CFPE), implicated in the progressive loss of lung function associated with CF lung disease. Over the course of a CFPE the common bacterial taxa show resistance to perturbations while the rare taxa show resilience. Through this investigation, *Veillonella parvula* was identified as a potential bioindicator of CFPE, introducing the potential for a rapidly testable parameter for clinicians to identify a CFPE. This finding could provide one of the most important recent developments in CF therapy.

Acknowledgements

Firstly I would like to express may sincerest thanks to all my supervisors; Dr Christopher van der Gast, Dr Anna Oliver, Dr Alan Walker and Dr Kenneth Bruce. Without the help and support you have provided over the last four years I could not have hoped to achieve so much. I would also like to make a special thanks to Dr Geraint Rogers, your help and advice throughout this project has been invaluable.

I would like to acknowledge the contribution of the clinicians and medical staff, around the world, who have worked tirelessly to provide me with samples and metadata required to produce this work. In particular I would like to thank Dr Mary Carroll and Professor Lucas Hoffman for their contributions to this work.

Thanks to all the past and present, staff and students at CEH Wallingford, in particular, Tom August, Rory O'Connor, Phil Martin, Will Hentley, Scott McKenzie, Carolin Schultz, Louise Barwell, Phill James, Lindsey Newbold, Dan Reid, and Bruce Thompson. It's been quite an adventure guys, thanks for all your help, support and friendship.

There are so many people out there who deserve thanks for their friendship and encouragement over the last few years. I wish I could thank you all, but in the interest of not writing another book I want to mention some key people. Please know that I am grateful to every single person who has been part of my life during this time.

A massive thank you goes to the amazing Helen Waters, you're a true friend and your dedication to everything you do is an inspiration. Tom Simpson and Sam Jones, you guys have been amazing, thank you for keeping life in perspective. Laura "weasel" Balfour, you always know when to show up when you are truly needed. Frankie Alcock, you are a star, couldn't have done it without you dude. I am so grateful to have you all as friends.

Finally, endless thanks go out to my wonderful family who have supported me tirelessly through everything I have set out to achieve. I would not be where I am today without you.

Contents

Table of Figures

Table of Tables

Chapter 1: General Introduction

General Introduction

“The role of the infinitely small in nature is infinitely great”

Louis Pasteur

1.1 Introduction

Lower respiratory tract infections (LRTI) are a considerable burden to human health and healthcare organisations. These infections represent a wide range of conditions, which include both, acute, short lived infections such as pneumonia and respiratory syncytial virus (RSV), as well as more chronic conditions, including chronic obstructive pulmonary disease (COPD) and the infections associated with cystic fibrosis (CF). The World Health Organisation (WHO) ranks LRTI as the leading cause of death due to infection (1) and as a result, the costs of LRTI economically, and in terms of morbidity and mortality, are immense (2).

Pulmonary failure as a result of progressive lung damage from chronic infection and concomitant inflammation is the primary cause of morbidity and mortality in CF patients (3). Given the influence of infection on clinical outcome, it is by extension important to characterise the species in the CF airways. In recent years advances in culture independent methodologies have revealed that CF lung infections are not, as previously considered, the result of a single or few pathogenic organism but rather comprise a complex and diverse microbial community (4). It is therefore important to fully understand the relationship between the microbial community and disease progression. To achieve this, changes in the bacterial community along with the factors that influence this, have to be investigated.

The aim of this dissertation was to investigate how bacterial community composition within the lungs of individuals with CF relates to disease state. By employing a multidisciplinary approach combining medical and ecological methodologies, it was possible to investigate the relationship between the bacterial community and disease state.

1.2 The microbial world

Microorganisms, including bacteria, archaea, viruses and fungi, are some of the most abundant forms of life on the planet. Current estimates suggest there are between 9.2-31.7x10^{29} bacterial and archaeal cells (5), about 10^{31} viral particles (6). Due to the multicellular nature of some fungal species, no approximation of the numbers of individual fungal organisms exists however, it is estimated that there are around 1.5x10^{6} species on Earth (7). Despite associations with health and disease, microorganisms are essential for maintaining the planet's ability to sustain life through essential roles in biogeochemical cycling (8). Concentrations of oxygen, carbon dioxide and methane in the atmosphere are influenced by microbes, which regulate functions such as the global hydrogen, carbon, nitrogen, oxygen and sulphur cycles (8).

Improvements to the microscope around 300 years ago opened up a new world of microorganisms. Robert Hooke, who published the first depiction of "microfungus" in 1665, stated "...in every little particle of its matter, we now behold almost as great a variety of Creatures, as we were able before to reckon up in the whole Universe itself."(9). Later in 1676, van Leeuwenhoek, using a microscope of his own design which magnified up to 250 times, visualised and described the first bacterial cells (10).

It was not till the 19th century that major leaps forward in the understanding of microorganisms and their role in disease were made. Robert Koch, considered by many to be the founder of modern medical bacteriology, laid out guidelines to establish the microbial cause of infectious diseases, known as Koch's postulates (11, 12). These guidelines state that:

1. The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms.
2. The microorganism must be isolated from a diseased organism and grown in pure culture.

3. The cultured microorganism should cause disease when introduced into a healthy organism.
4. The microorganism must be reisolated from the inoculated, disease experimental host and identified as being identical to the original specific causative agent.

To do this he developed methods for the pure culture of bacterial species, which allowed microbes to be classified into taxonomic groups by appearance and nutritional requirements. These methodologies are still widely used in diagnostic microbiology, however, with advances in diagnostic methodologies modification to the original Koch's postulates have been suggested by Fredricks and Relman (1996) (13);

1. A nucleic acid sequence belonging to a putative pathogen should be present in most cases of an infectious disease. Microbial nucleic acids should be found preferentially in those organs or gross anatomic sites known to be diseased, and not in those organs that lack pathology.
2. Fewer, or no copies of pathogen-associated nucleic acid sequences should occur in hosts or tissues without disease.
3. With resolution of disease, the copy number of pathogen-associated nucleic acid sequences should decrease or become undetectable. With clinical relapse, the opposite should occur.
4. When sequence detectiomn predates disease, or sequence copy number correlates with severity of disease or pathology, the sequence-disease association is more likely to be a causal relationship.
5. The nature of the microorganism inferred from the available sequence should be consistent with the known biological characteristics of that group of organisms.
6. Tissue-sequence correlates should be sought at the cellular level: efforts should be made to demonstrate specific in situ hybridisation of microbial sequence to areas of tissue pathology and to visible microorganisms or to areas where microorganisms are presumed to be located.

7. These sequence-based forms of evidence for microbial causation should be reproducible.

While the modified poslulates are useful they do not account for established disease association, such as papillomavirus and cervical cancer or prion disease (13).

DNA was first discovered in 1869 by Freidrich Miescher as a microscopic substance located in the nuclei of cells. Watson and Crick's work on the structure and function of DNA (14) and the subsequent understanding of the role of DNA in an organisms evolutionary history (15), led to Carl Woese proposing the use of ribosomal RNA (rRNA) sequences to classify the taxonomy of microorganisms (16). The use of rRNA for the investigation of microbes began to reveal the enormous genetic diversity present in the microbial world (17). As work continued in this area it became apparent that only a fraction of the bacteria from the environment had been grown in the laboratory (18).

The vast genetic diversity of microorganism is still being revealed. The *Approved Lists of Bacterial Names* (http://www.bacterio.net/-number.html) as of August 2013, included 10,599 published species names, a more than 500% increase from those listed in 1980. This increase in recognised taxa is attributed largely to advances in DNA sequencing technologies.

The 16S rRNA gene codes for part of the ribosomal RNA which makes up the 30S prokaryotic ribosomal subunit. The S in 16S refers to the Svedberg unit, which is a non-SI unit for sedimentation rate. The 16S rRNA gene is the most commonly sequenced section of the bacterial (or archaeal) genome used for identification (19). It is around 1,500 bp in length and contains highly conserved regions of the rRNA gene, present within different species of bacteria and archaea, and has nine hypervariable regions (20, 21). The conserved regions allow the design of universal primers that will amplify "all" bacterial or archaeal species while the hypervariable regions that can provide species-specific signature sequences important in allowing species identification (20, 21).

For similar reasons, the internal transcribed spacer (ITS) regions of the rRNA gene are the most widely used for the identification of fungal species (22). The ITS regions are lengths of non-functional RNA, interspersed along the conserved rRNA gene (18S_ITS1_5.8S_ITS2_28S). These areas have been widely sequenced due to their high copy number, allowing amplification from small quantities of DNA, and their high variability compared to other genetic regions of ribosomal DNA, thereby allowing identification of even closely related fungi to the species level (23).

Unfortunately, no universal primer equivalent exists for the molecular identification of viruses due mainly to the variation in genetic material, single stranded DNA, single stranded RNA, double stranded DNA or double stranded RNA. Therefore the diagnosis of viral infections has relied heavily on the use of multiplex PCR, developed for panels of known viruses (24). In recent years however, methods have been developed using metagenomics, viral particle purification followed by shotgun sequencing, to identify unknown and previously uncharacterised viral particles from a variety of sample types (25).

1.3 Microbial ecology

Despite the diversity, ubiquity and importance of microorganisms, microbial communities are still poorly understood. The discipline of microbial ecology aims to address this by investigating microbes and their interactions with the environment. This is arguably one of the most important challenges for ecologists today, as understanding these communities may have major impacts on some of the fundamental trials facing society, such as the management of natural ecosystems and the mitigation of climate change (26).

Ecological theory is essential in order to interpret and predict interactions between species in a shared environment (27). However, the abundance, diversity and actions of microorganisms make it challenging to apply theories developed for the study of plants and animals to microbial communities. This is mainly due to the challenges associated with

assigning microorganisms to species. The traditional definition of species involves the ability to produce viable offspring from sexual reproduction, the biological species concept (28). Therefore, this cannot be applied to organisms that reproduce asexually, and hence in order to apply ecological theory to microorganisms it became important to determine a new method to differentiate between microbial species, this is explored further in Chapter 1.4.

Despite some challenges in applying ecological theory to microbial populations, ecology has been able to generate predications of practical value for social, political and business needs. An example of this included the application of epidemiological models in predicting the spread of pathogens (29). Quantitative theory has also been employed to optimise operating conditions for the efficienct running of wastewater treatment plants (30). On a wider scale it is hoped that quantitative information on microbial community structure and population dynamics will help assess the impact of climate change on microbial ecosystem processes.

The challenge facing microbial ecologists is to match the appropriate theoretical approach to the organism, system, scale and question of interest. Until recent years lack of available information about the microbial community was a limiting factor of these studies, however the development of high throughput sequencing platforms has lead to greater use of ecological theories.

In the inverstigation of human microbiota, each individual can be considered an island habitat, containing a diverse microbial assembalage (31). As a result ecological theories such as the community assembly theory (32) and metacommunity theory (33) can be utalised to help understand the community dynamica and variability in the human microbial community (31).

1.4 Identifying bacteria

Traditionally, microbes were clustered using phenotypic characteristics, subsequently these methods were soon overtaken by the use of genotypic approaches. Whole genome methods such as DNA-DNA hybridisation (DDH), currently considered the gold standard for species

identification, and average nucleotide identity (ANI), a more recent technique that uses a pair wise comparison of all shared sequences between two strains, are widely used. However, these methods are time consuming and technically demanding, compared to rapid and easy sequencing of signature sequences such as the 16S rRNA genes (34, 35). It is not surprising therefore that 16S rRNA amplicons are commonly sequenced to investigate the phylogeny and taxonomy of bacterial species (36). When comparing full length 16S rRNA gene sequencing, two microorganisms are defined as the same species if their ribosomal RNA gene sequences show over 98.7% similarity (37).

The 98.7% cut off is considered appropriate for bacterial species identification as it accounts for a certain amount of within species variation. Variation is observed within nine hypervariable regions (V1-V9) found interspersed with conserved areas of the 16S rRNA gene. As sequencing of the entire 16S rRNA gene is not always practical, smaller regions of the gene may also be used for species identification. As such, the V1-V2 or the V3 to V5 regions of the gene can be used for the taxonomic classification of bacterial sequences. By covering more than one hypervariable region, sensitivity, specificity and reliability are increased. Even so, due to reduction in the conserved areas being covered in these shorter reads, a 97% similarity cut off can be used as an approximation for species.

Our ability to amplify and sequence rRNA genes is important in enabling species diversity to be estimated. This approach can have limitations; a study by Hong *et al*, 2009, using clone libraries of marine sediment, found that polymerase chain reaction (PCR) of 16S rRNA genes with ‘universal’ primers failed to amplify at least half of the microbial diversity regardless of taxonomic level (38). Further, Gans *et al* (2005) calculated that in order to estimate 80% of the diversity within a soil sample, containing an estimated billion bacterial cells, at least 1 million 16S rRNA gene reads would be required (39). Although, this may have previously seemed unachievable, recent technological improvements are now making such assessments feasible, as reviewed by Shendure and Aiden, 2012 (40).

DNA profiling methodologies such as Terminal Restriction Fragment Length Polymorphism (T-RFLP) have, and continue to, allow microbial ecologists to monitor changes in the structure and composition of microbial communities. T-RFLP is one of the most frequently used profiling techniques for culture independent studies of bacteria and fungi due to its speed, simplicity and reproducibility (41). T-RFLP is based on PCR amplification of a target gene with fluorescent labelled primers; amplicons are then digested with restriction enzymes to create fragments, which can be separated and detected using a DNA sequencer, thereby creating a profile of the community. These profiles are shown graphically as peaks, where the X-axis represents the fragment size and the Y-axis, the fluorescence intensity. Each peak is associated with a genetic variant within the original sample, while the intensity indicates the relative abundance within the community. The profiles created can be used to compare communities, by peak presence or absence, while clone libraries or peak resolving databases can be used to identify the species associated with each peak. This process was carried out by Rogers *et al* 2003, for the investigation of the bacterial community within the CF lung (42). Despite the frequency of use, T-RFLP has been criticised for its lack of resolution (43, 44) and therefore in recent years there has been a drive towards the use of high throughput or next generation sequencing technologies (45).

The move towards next generation sequencing has not been without challenges, particularly as early technologies were hampered by the short read length produced, limiting taxonomic resolution (46). However, as technologies have improved so have read lengths, and therefore the use of next generation sequencing has increased exponentially. Technologies such as Roche 454 (47) or Illumina MiSeq (48) are now widely used and capable of producing hundreds of thousands of sequence reads per sample. These can be directly clustered into operational taxonomic units (OTU, an operational definition of species used in the analysis of DNA sequencing data) thereby allowing better estimates of community diversity to be calculated.

1.4.1 Human microbiota

Humans are superorganisms composed of both human and microbial components (49). As more is understood about the microbiota associated with the human body, the importance of microbes in human health is being increasingly recognised. In fact, it is estimated that the microbes associated with the human body outnumber our cells by 10-fold (50).

Since the 19th century, culture based analysis has been the principle method for the investigation of human associated microbes (51). However, with the vast number of microbes present on and within the human body, the use of culture for the investigation of these dense and complex microbial communities can be problematic (50). The introduction of DNA based analyses, in particular next generation sequencing technologies, has given us the ability to create vast quantities of data that can be used to investigate the composition and function of microbial communities associated with the humans.

In 2008, The International Human Microbiome Consortium (IHMC) was officially launched, with the aim to study and understand the role of human associated microbes in the preservation of health and the causation of disease (http://www.human-microbiome.org). This work has revealed a complex community of microorganisms, where community composition primarily varies by anatomical site (51). Understanding the complex microbial community associated with the healthy individual is the first step in understanding the complex relationship between the microbial population and human health.

To fully understand the relationship between the microbial community and human health understanding colonisation and succession is key. Inside the amniontic sac the faetus is considered essentially sterile, however, after vaginal delivery the microbial community is found to closely resemble that of the mother's vagina (52). The predominant organisms in the vagina are Lactobacilli, therefore it is unsurprising that this is the first organism found in the gastrointestinal (GI) tract of newborn babies (52). This initial colonisation is considered to prepare the GI tract for further microbial colonisation and succession (52). Over time a

densely populated microbial community develops over the body, with changes occurring due to changes in the body, the eruption of teeth (53), as well as external environmental factors. Cho *et al* (2012) has been hypothesised that changes in the microbiotia in early life can affect host immunological, metabolic, cognative and reproductive development (52).

A wide variety of studies have been published examining the link between microbial populations and disease states. These have revealed changes in the microbial community associated with, for example; psoriasis (54), obesity (55), asthma (56), colitis (57) and cardiovascular disease (58). The challenge for these studies is to identify whether there is a causal association in the microbial variation.

Through the use of metagenomics, and metaproteomics the human Microbiome Project (HMP) is aiming to more fully understand the links between the microbial community and human health and disease (59).

1.5 Challenges of the respiratory tract

The respiratory system, shown in Figure 1.1, is designed for the efficient transportation of oxygen from the environment to the blood stream and the removal of carbon dioxide (60). In detail, air enters the body through the mouth or nose and travels down through the pharynx, into the trachea. The trachea then splits into the two bronchi, which subsequently branch into the bronchioles and terminate with alveoli. The alveoli are the site of gas exchange; oxygen is transported across the epithelial layer into the bloodstream to be transported to cells for respiration, while carbon dioxide is removed and travels back out the lungs to be expelled from the body (60). As such the lungs are the only interior organ constantly exposed to the exterior environment and are therefore vulnerable to any particles or organisms present in the air.

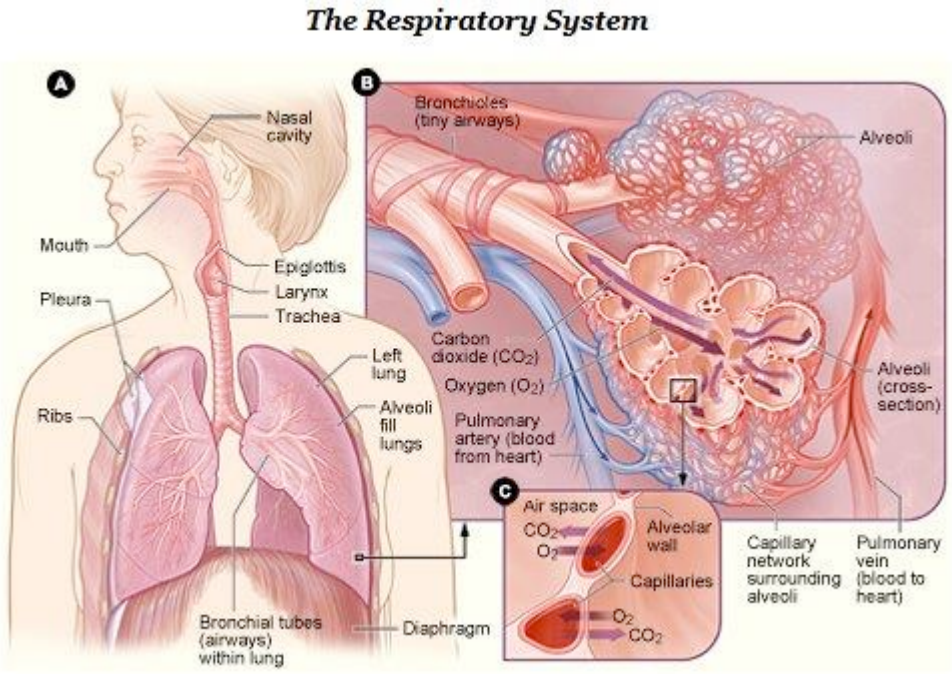


Figure 1.1 Schematic diagram of the respiratory system.

A) The respiratory system within the human body, B) an enlarged view of the alveoli and surrounding capillaries, C) schematic indication the movement of gas between the alveoli and the blood stream.

It is estimated that the air in the Earth's atmosphere contains around 1000 million tonnes of suspended particles (61). These particles range in size from 0.1-1000μm and include pollen, smoke, soot, dust and importantly microorganisms (62, 63). In 2007, Brodie *et al* used molecular techniques, described in detail later, to analyse the bacteria present in aerosols from urban areas of San Antonio and Austin (USA) over a 17 week period (64). In this study over 1,800 bacterial taxa were identified, including potentially pathogenic species such as, *Burkholderia mallei* (Glanders disease), *Burkholderia pseudomallei* (melioidosis) and *Clostridium botulinum* (botulism) (64).

Whilst there will be some variation due to location and environmental factors (65), inside a building it is estimated there are between 400 and 900 individual microorganisms per cubic metre of air (62). Assuming an adult inhales around six litres of air per minute, this would mean that *c.* 10,000 microorganisms would be inhaled per day (61). Although the vast majority of these organisms are likely to be non-pathogenic, the risk of encountering pathogenic organisms is significantly increased in clinical environments (48). This is particularly true in dental surgeries where high-speed vibrating instruments have been found to aerosolise oral bacteria such as Staphylococci and members of the viridans streptococci group, the causative agents in many mouth or gingival infections as well as tooth decay (66).

The risk of infection by the inhalation of airborne organisms is reduced further by the ability of the healthy airways to efficiently remove inhaled particles (61). As the air passes though the nasal cavity particles are filtered out, with between 70-80% of the particles that are 3-5 μm in diameter and 60% of those around 2μm, being trapped in the hair present in the nostrils (60).

With the exception of the terminal bronchioles and the alveoli, the rest of the respiratory tract is lined with ciliated cells, mucus secreting goblet cells and subepithelial mucus-secreting glands (60). The ciliated epithelium is also covered by a two-layer film known as the airway surface liquid (ASL), made up of a lower layer known as the periciliary fluid and an upper mucus gel layer (67). Particles or microorganisms breathed into the lungs become trapped in

the mucosal secretions and are transported by the beating cilia out of the lungs to the back of the throat where they can be swallowed or expectorated, this system has been described as a mucociliary escalator (60). It is estimated that this escalator system can move entrapped particles or organisms from the airways at a rate of 1 cm/minute (61).

If particles or microorganisms make it past the mucociliary surfaces of the nasal cavity and the lung, the alveoli are lined with antimicrobial peptides and host defence cells, such as macrophages. While macrophages have the ability to phagocytose any particle or organism that make it past the mucociliary escalator (61), antimicrobial peptides work as natural antibiotics, protecting the lungs from invading microbes. Antimicrobial peptides, such as β-defensins and lactoferrins, are produced by epithelial cells, macrophages, neutrophils and natural killer cells and act as natural antimicrobials against any organisms entering the respiratory tract (68, 69). As a result it is very difficult for microorganisms to establish infections in healthy individuals (70).

It is well established that the upper airways (nasal cavity, pharynx and larynx) are colonised with a wide range of microorganisms which make up the “normal” microbiota (71). However, until recently it was thought that the lungs were a sterile environment, with no airborne microorganisms surviving the innate immune defences (72, 73). Work by Charlson *et al* 2011, using 454 16S rRNA pyrosequencing to analyse the bacterial community from six healthy patients, established that the lungs of healthy individuals contains similar sequences to those observed in the upper respiratory tract however, in lower abundance (72). Although results published differ, the most commonly identified genus include; *Pseudomonas, Streptococcus, Prevotella, Fusobacteria,* and *Veillonella* (74-76). Differences observed may be the result of the different methodologies employed by these studies or the geographic locations at which they were carried out.

1.6 Lower respiratory tract infections

The WHO lists LRTIs 3rd on the top 10 causes of death factsheet (1), with around 6% of the world's total mortality being a result of these infections. This translates to 3.8% of deaths in high income countries and as much as 11.3% of deaths in countries with lower income. Importantly these figures do not include tuberculosis, which itself is ranked as the 8th biggest cause of death worldwide (1).

As explained previously there are many forms of LRTI (77), as a result of this variety and high incidence of these infections, they represent a huge burden to national healthcare systems. For example, around 4% of the total National Health Service (NHS) budget is spent on problems associated with the respiratory system, amounting to around £4.43 billion (2). It is therefore both clinically and economically important to investigate better methods to prevent and treat LRTIs (78, 79).

To achieve this however, we have to understand the underlying causes of disease. In recent years advances in culture independent technologies has led to advances in our understanding of human associated microbial communities in general.

1.7 Cystic fibrosis

Wehe dem Kind, das beim Kuß auf die Stirn salzig schmekt, er ist

verhext und muss bald sterbe

"Woe to the child who tastes salty from a kiss on the brow, for he is cursed and soon must die" 18th century German literature (80).

CF is the most common recessively inherited genetic disorder primarily affecting the Caucasian population, although it is seen in low frequencies in other populations; a study

carried out in 2011 revealed that 1 in 59 Hispanic, 1 in 84 African American and 1 in 242 Asians were carriers of CF mutations (81).

There are approximately 70,000 people suffering from CF worldwide, with around 30,000 individuals living in the USA (82) and over 9,300 in the UK (83). Around 1 in 25 people in the UK carry CF causing mutations however, these numbers vary across the world (84).

CF is a multi-systemic disorder, which has major effects on a number of body systems notably the gastrointestinal tract, the reproductive tract and the focus of this investigation, the respiratory tract. Despite medical advances, between 80 and 95% of CF patients die as a result of chronic lung infections and concomitant inflammation (85). Once infection has established, eradication of infecting species can be extremely difficult (86, 87). As a result of these infections and the inflammatory immune response they initiate, there is an irreversible loss of lung function over time, which leads to the majority of CF patients dying from respiratory failure (88).

1.8 Cystic fibrosis history

Although CF has been observed for centuries, it was not until 1938 that Dr Dorothy Andersen, a pathologist at the New York Babies Hospital, described the condition fully (89). Prior to this study of malnourished infants, CF was misdiagnosed as coeliac disease and resulted in the death of many children before they reached 6 months old. Andersen's autopsy study of infants who died of this condition revealed it to be a distinct disease, termed "cystic fibrosis of the pancreas".

The disease was characterised by mucus plugging of the glandular ducts, a decreased ability to absorb fat and proteins, steatorrhoea (the presence of excess fat in faeces), failure to thrive and pulmonary infection (89). Rather than an intolerance to gluten, nutritional failure

was thought to be due to damage to the pancreas and lack of pancreatic enzymes, leading to a vulnerability to lung infections, often noted as the cause of death (89).

In 1946, CF was recognised as an autosomal recessively inherited genetic disease (90). At this time, CF research focused on mucus abnormalities as the cause of CF. It was a further two years before Paul di Sant'Agnese observed that the levels of sodium and chloride ions in the sweat of patients diagnosed with CF were five times higher than that of healthy individuals thereby implying the mucus glands were not, as previously thought, the site of the basic defect (91, 92). The discovery of the ion abnormalities in the sweat of CF individuals lead to Gibson and Cooke developing the "sweat test" (93). This test was found to be highly successful due to its discriminating power, allowing easy identification of patients, even those with milder symptoms.

It was not until the 1980's that major advances in the understanding of CF came about. In 1983 it was identified that abnormalities in ion transport, resulting in low absorption of chloride ions and an increased reabsorption of sodium ions, were responsible for CF (94). The mutated gene responsible for CF was discovered in 1989 (95-97). The gene was found to code for a cyclic-AMP-regulated chloride transport channel, named the cystic fibrosis transmembrane conductance regulator (CFTR), normally expressed in epithelial cells (95-97).

1.9 The CF mutation

The CFTR gene, located on the long arm of chromosome seven (95), codes for a membrane bound ATP binding cassette (ABC) transporter protein. This protein is found in the apical membrane of epithelial cells (84) and maintains the osmotic balance across epithelial cells by controlling the transport of chloride ions (98). Mutations in both copies of the CFTR gene result in CF disease symptoms. These mutations translate to a wide range of physical manifestations including; infertility due to obstructive azoospermia, which effects 98% of CF

males (99); pancreatic insufficiency, leading to fat malabsorption (100); liver cirrhosis, as a result of ductular obstruction and abnormal bile flow (101); diabetes, as a result of diminished insulin production (102); and chronic inflammatory lung disease (84). However, there are also cases where mutations are not associated with any CF symptoms (103).

Currently more than 1800 different mutations of the CFTR gene associated with CF have been identified (103). These mutations have a range of effects on the functioning of the CFTR and can be categorised into five different classes, resulting in a range of disease severities, illustrated in Figure 1.2 (104). “Classic” (class I or II) CF is characterised by a loss of function in both alleles resulting in no functioning CFTR. This is associated with pancreatic insufficiency and progressive lung disease (105). In some cases the mutation in one allele can result in a partially functioning CFTR associated with an improved prognosis (106). Even so, although some information on disease severity can be gained from information on the gene mutation, not all patients with the same mutation experience the same disease outcome (103, 106).

The most prevalent mutation resulting in CF is the result of a deletion of phenylalanine at position 508 of the gene (ΔF508). It is estimated that approximately 70% of all CFTR mutations are due to the ΔF508 mutation (92) however, this number varies with geographic location as shown in Figure 1.3 (107). ΔF508 is a class II mutation resulting in misfolding of the CFTR protein which is then degraded in the endoplasmic reticulum and golgi body, as a result no functional CFTR protein is present in epithelial cells (108).

The identification of the basic genetic defect resulting in CF has lead to the hope that gene therapy, inserting a functional copy of the CFTR gene into appropriate cells, could be the “key” to curing CF (92). Currently the UK Cystic Fibrosis Gene Therapy Consortium is carrying out five clinical trials using cationic lipids to carry genes into the respiratory tract of CF individuals. However, at this time only partial correction of the CFTR defect has been achieved and results still require improvement (http://www.cfgenetherapy.org.uk/).

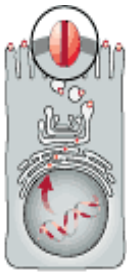
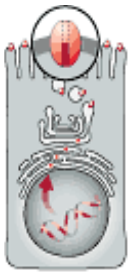
Defect Classification	Normal	I	II	III	IV	V
						
Defect Result		No synthesis	Block in Processing	Block in Regulation	Altered Conductance	Reduced Synthesis
Types of Mutation		Nonsense; Frame shift	Missense; Amino Acid Deletion (ΔF508)	Missense; Amino Acid Change (G551D)	Missense; Amino Acid Change (R117H) (R347P)	Missense; Amino Acid Change (A445E) Alternative Splicing

Figure 1.2 CFTR mutations associated with CF have been classified into five broad classes based on their effect on the CFTR transporter protein.

Adapted from the original by Zielenski and Tsui, 1995 (104), illustrations reprinted from the image, Ratjen, F, 2007, New pulmonary therapies for cystic fibrosis, Current Opinion in Pulmonary Medicine, **13**:541-546 (109), under copyright licence from Lippincott Williams & Wilkins, Inc, License Number: 3471420161666.

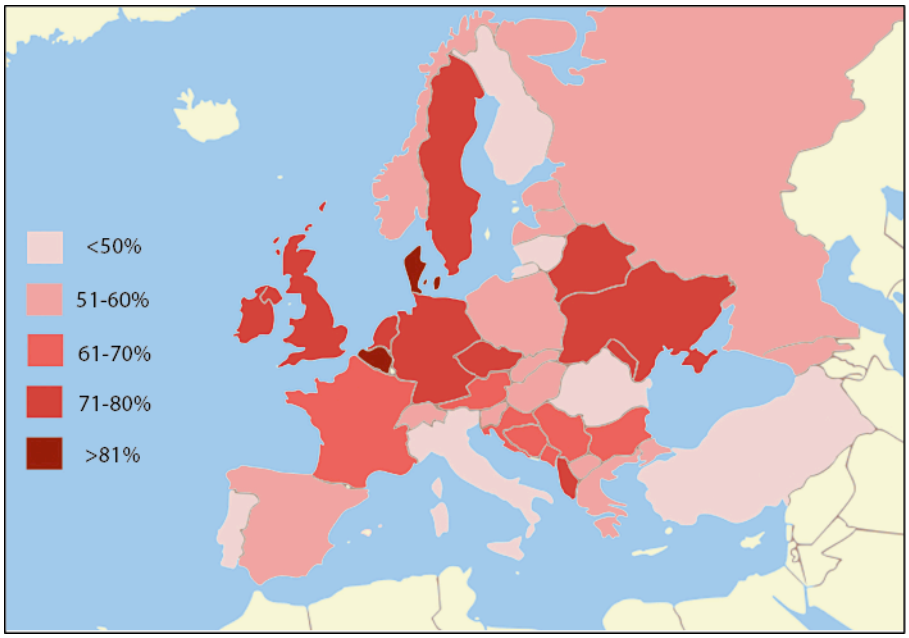


Figure 1.3 Geographical representation of the prevalence of the CFTR mutation ΔF508 as a percentage of all CFTR mutations and recognised by country.

The prevalence of ΔF508 is related to the incidence of CF, countries with higher CF rates have greater number of CF cases. Prevalence was found to be greater in more northerly countries, while prevalence is generally low in Mediterranean countries.

The figure is re-used in unmodified form from, Southern, 2007, Cystic Fibrosis and Formes Frustes of CFTR-Related Disease, Respiration **74**:241-251(107), by permission of Karger Publishers, licence number 3471430204165.

1.10 The CF lung

1.10.1 Mucus

Despite the multisystemic nature of CF, the maintenance of airway function is particularly important as between 80 and 95% of CF suffers ultimately die as a result of chronic infection and associated airway inflammation (85). Improper functioning of the CFTR protein within the lung leads to an inability to clear microorganisms and results typically in chronic lung infections. Two main theories have been suggested as to how this happens; the first is known as the 'high salt' hypothesis. This theory suggests that while the healthy lung is protected against infection by antimicrobial peptides, in the CF lung the absence of a functional CFTR protein leads to a build up in the concentration of NaCl on the airway surface, which degrades these antimicrobial peptides (110).

In contrast, the second hypothesis, known as the 'low volume' hypothesis, explains the chronic infections in CF by suggesting that there is too little salt on airway surfaces leading to airway surface dehydration. This in turn, leads to thickened mucosal secretions, breakdown of the mucociliary escalator and an inability to clear invading organisms from the lung (111).

The 'low volume' hypothesis is the more widely accepted explanation for the development of lung disease. In more detail, the defective Cl^- ion secretion, as a result of CFTR mutations, leads to epithelial Na^+ ion channel hyperactivity, resulting in the osmotic movement of water out of the ASL (112). As the ASL becomes dehydrated, the cilia are compressed, unable to beat and move the viscous mucus from the lower respiratory tract. When the periciliary fluid becomes critically dehydrated mucins within the mucus layer bind to those anchored to the epithelial cells. As mucociliary clearance has broken down at this stage the airways become vulnerable to infection (111). The failure to clear mucus from the airways, alongside the continual production of mucins by goblet cells and glands results in airway surface plaques and intraluminal plugs which ultimately lead to airway obstruction and provide the main site of infection (67).

1.10.2 Inflammation

As well as being characterised by airway obstruction and persistent infection, a vigorous inflammatory response plays a key role in the progression of CF lung disease. Epithelial cells orchestrate the inflammatory response by producing an abundance of IL-8 (Interleukin 8) and other proinflammatory cytokines which induce a neutrophil response (113). Unfortunately, the over stimulation of neutrophil activity is combined with the deficiencies in immunoregulatory factors, IL-10 and nitric oxide, resulting in an excessive and prolonged inflammatory response (113). This in turn leads to lung tissue damage, which over time contributes to loss of lung function (114).

Inflammation is a means by which the body is able to contain and/or control infection. However, within the CF lung this mechanism breaks down, becoming a pathological force. As stated above, it is unclear if inflammation in early infection is triggered by invading organisms, although it has been well established that the magnitude of the inflammatory response is correlated with bacterial burden (115).

Evidence suggests that with the exception of some of the sub mucosal gland ducts, infants with CF are born with normally developed lungs. Never the less, shortly after birth pulmonary symptoms start to establish, with infection and inflammation being observed soon after diagnosis by newborn screening. Even so, it is still unclear if the inflammatory response is triggered by infection or is already a factor of the CF lung (116). What is clear however is that airway inflammation in early life is associated with reduced lung function and structural lung damage (117).

1.11 CF lung infections

In childhood years, infection tends to be transient with chronic infections establishing over time, leading to progressive and irreversible lung damage (85). Traditionally, CF lung

infections have been considered to be the result of either a single or small number of species (118). With further investigation however, it is becoming clear that this is not representative of the true number of organisms present within the CF lung, which may be considered a complicated and diverse ecosystem (119).

Over time the CF lung is under constant pressure from a range of factors, including, viruses, bacteria, fungi, allergens and irritants, creating a dynamic perturbed system (120). Perturbations within the lung ecosystem can lead to acute respiratory symptoms, defined as CF pulmonary exacerbations (CFPE) (121). Despite the importance of CFPE in CF disease progression no formal definition or standardised criteria exist to define them. Instead clinical symptoms such as; increased cough, loss of appetite, weight loss and decline in lung function, are used by clinicians to define the start and end of CFPE periods (121).

In order to help maintain lung function antibiotics treatments are continually administered and manipulated by clinicians. During periods of CFPE clinicians intervene, generally with administration of intravenous antibiotic therapies (121). Understanding how the microbial community changes leading into, during and after exacerbation may give insight into how the treatment can be used to maintain lung function and slow disease progression.

1.11.1 Sampling

Antimicrobial therapies are an important factor in preserving lung function. It is therefore vital to investigate the microbial community within the CF lung. A number of sampling methods are recognised for the analysis of the microbiology associated with CF lung infections. These include; cough swab, cough plate, oropharyngeal culture, laryngeal or nasopharyngeal aspirate, expectorated sputum, induced sputum, bronchoalveolar lavage (BAL) and bronchoscopy brush specimens (122). While all these methods are used, the most commonly used sampling methods are, expectorated sputum, induced sputum and BAL.

All sampling methods introduce the risk of sampling error or bias. Sampling error occurs due to observing only a subset of the population rather than the whole population. It was

therefore important to consider the pros and cons of all sampling methods available before the sampling method was chosen.

BAL is considered the gold standard for assessing the microbial community within the CF lung, particularly for young children who do not produce sputum, as it has been shown that non-BAL samples do not adequately reflect the bacterial community present in the lower airways in these individuals (123, 124). However, it is an invasive procedure that involves using a bronchoscope to infuse sterile saline into the lungs. The saline is then suctioned out of the lungs into a collection trap, this fluid can then be analysed to assess the microbiology of the lung (104). While this method has been found to be effective, its highly invasive nature requires patients to be anesthetised (124, 125).

As a result the CF Trust standard laboratory procedures recommend the use of sputum for routine sampling. Sputum is used as an important sampling tool for the investigation of chronic lung disease, as it allows the investigation of protein, cellular and importantly, microbiological components which effect disease progression and severity (126). There are two methods for obtaining sputum samples, spontaneous expectoration and induction.

Spontaneously expectorated sputum is the most widely used sampling method for CF adults and has been found to provide comparable microbiology results to other sampling methods, including induced sputum and BAL, using both culture based and molecular comparisons (127-129). However, not all patients are able to spontaneously produce sputum, particularly those with mild disease symptoms. In these cases sputum induction may be used. Sputum can be induced by nebulising isotonic or hypertonic saline, and has been shown to be a simple, non-invasive and cost effective sampling procedure (128, 130). A study by Henig *et al* (2001) revealed that induced sputum tended to show higher identification of recognised pathogens than BAL, while being preferred by patients, making it an important alternative to BAL for older children and non-sputum producing adults (130).

This project focused on the bacterial community present in the lung of CF adults. Consequently, throughout this study spontaneously expectorated sputum was used as the sampling method. Spontaneously expectorated sputum was chosen as the least invasive procedure which has also been shown to provide a comparable sample of the bacterial community compared with other sampling techniques.

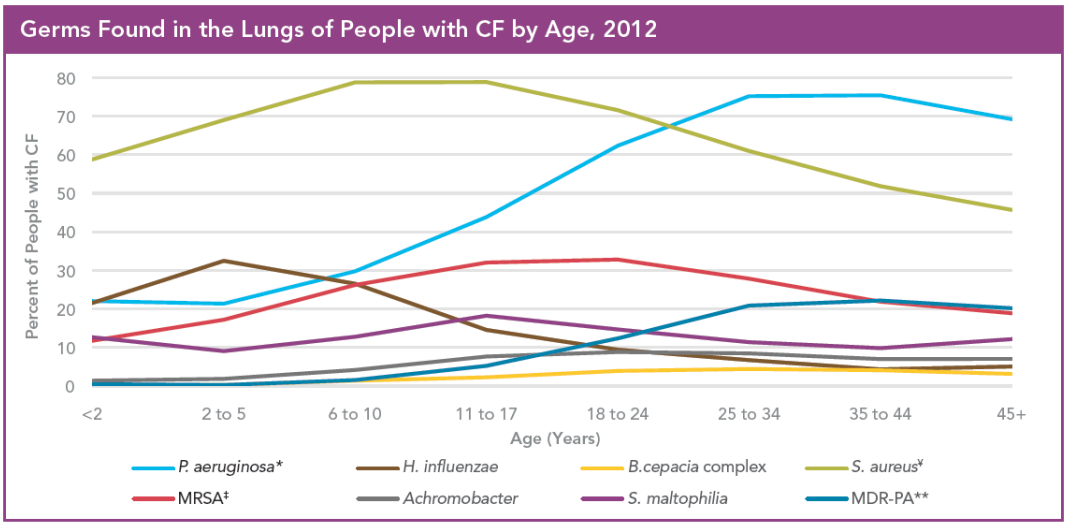
1.11.2 Traditional microbiology

Routine diagnostic analysis of CF respiratory samples has traditionally involved selective culture, using a variety of selective media and growth temperatures, in aerobic conditions, to quantify and identify a number of known organisms, (131). This is generally confined to a small targeted group of microbes that are considered clinically important, including *Staphylococcus aureus*, *Haemophilus influenzae, Pseudomonas aeruginosa, Stenotrophomonas maltophilia, Achromobacter xylosoxidans*, *Burkholderia cepacia* complex (BCC) (3), as well as species of the filamentous fungi *Aspergillus* (most frequently *A. fumigatus*) and yeast (commonly *Candida albicans*) (132, 133). Although all these organisms are considered to play a role in disease progression, *P. aeruginosa* has been regarded by many to be the most significant pathogen in CF due to its prevalence and persistence within the lung (134).

Diagnostic microbiology has been useful in revealing key pathogens involved in the establishment of chronic infections and the subsequent loss of lung function. Supporting the understanding that over time there is a sequential acquisition of bacterial pathogens, Figure 1.4 (135). This model suggests that early infection is dominated by *S. aureus*, with other pathogens present in a much lower percentage of the population. However, in later years an increased colonisation by *P. aeruginosa* is observed. This soon becomes the most dominant pathogen, colonising the lungs of around 75% of CF sufferers (135).

Although it is well understood that these “recognised pathogens” play a role in the progression of CF lung disease, more evidence is being gathered suggesting that our

understanding of the microbial community is incomplete. For example, the use of non-selective culture media has indicated the presence of a diverse community of organisms not traditionally considered CF pathogens (infectious microorganisms considered to be the cause of respiratory symptoms in individuals with CF) (118). Furthermore, the discovery of a sharp oxygen gradient in CF lung mucus and subsequent anaerobic *in vitro* culturing of sputa has lead to the detection of still more bacterial species (132). Despite much effort, culture based methods have been shown to be limited (43). There are many reasons for this, perhaps most important is the likely under representation of microbial diversity due to the unknown laboratory growth requirements of bacterial and fungal species that may be present within the lung ecosystem (136). These requirements include; chemicals (Carbon, NitrogenSulphur, Phosphorus, Calcium), pH, Oxygen (aerobes, facultative aerobes, obligate anaerobes, Aerotolerant anaerobes, Microaerophiles), intracellular requirements and synergistic relationships with other organisms.



*P. aeruginosa *includes people with MDR-PA.*
***MDR-PA is multi-drug resistant* Pseudomonas aeruginosa (P. aeruginosa).*
*¥ S. aureus *includes people with MRSA.*
‡ MRSA is methicillin-resistant Staphylococcus aureus (S. aureus).*

Figure 1.4 Age dependent bacterial colonisation of the CF lung.

Reprinted from the 2012 Cystic Fibrosis Foundation annual data report (135). Data was sourced from Cystic fibrosis patients under care at CF Foundation-accredited care centers in the United States, who consented to have their data entered in 2012.

1.11.3 Culture independent technologies

Molecular techniques are revolutionising the way diagnostic tests and research is being carried out in relation to CF lung infections (43). In order to effectively investigate the microbial community using molecular methodologies, a representative sample of the nucleic acid from the entire microbial community is required. DNA extraction methods are an important route for the introduction of bias when carrying out DNA based analysis. It is therefore vital to select a method that minimises these effects, equally recovering DNA from gram positive, gram negative and fungal cells, providing a complete picture of the community (137). The methods employed throughout this project was modified by Rogers, *et al*, 2003, from a method described in 1989 by Pitcher *et al* (43, 138). This method employs physical disruption, in the form of bead beating, in conjunction with chemical guanidium thiocyanate-EDTA-sarkosyl (GES) lysis. Both physical and chemical disruption were included to disrupt the maximum number of cells and cell types and gain an accurate representation of the microbial community, for full methods see Chapter 2.3.

Techniques such as, PCR, T-RFLP and 16S rRNA gene sequencing (Sanger sequencing) have provided fresh insight in the analysis of both bacterial (43) and fungal (139) communities within the CF lung. Rogers *et al*, 2003, showed it was possible to rapidly characterise the bacterial community composition and diversity using 16S rRNA gene T-RFLP. In 2010, Nagano *et al*, used a mixture of non-selective and selective fungal culture to provide a comparison to results obtained from direct DNA extraction and sequencing, demonstrating a diverse fungal community that is potentially underestimated by mycological culture alone. They also demonstrated that potentially important pathogenic fungi may be missed by culture alone (139).

As technologies have moved on, more in depth analytical techniques for sequencing DNA; 454 high throughput pyrosequencing (140), Illumina sequencing (141), have allowed increased insight into the microbial communities within the CF lung by providing large quantities of data on the community within a sample. Roche 454 technology works using

emulsion PCR and pyrosequencing. Sequences are amplified onto beads in an emulsion PCR before being washed onto a PicoTiterPlate, where sequences attached to a single bead are added to each well. Bases are washed over the PicoTiterPlate in order, as a base binds a phosphopyruvate molecule is released, which is used to change adenosine 5´ phosphosulfate (APS) to Adenosine triphosphate (ATP) in the presence of sulphurylase. This releases ATP, then used to change luciferin to oxyluciferin in the presence of luciferase, releasing light. On the other hand Illumina Miseq is based on Solexa sequencing by synthesis chemistry, it uses a reversible chain termination method where a flouraphor bound to each nucleotide acts as a chain terminator. All bases can be added simultaneously, once bound lasers with filters are used to take an image for each base, before the flouraphor is cleaved and the new bases can be added.

By comparing the use of 454 and Illumina sequencing technologies Maughan *et al* (2012) revealed that Illumina sequencing provided better taxonomic resolution than 454 (121). In comparisions of benchtop sequencers, Loman *et al* (2012) and Junemann *et al* (2013) found that 454 provided the longest read lengths, however, Miseq provided the highest throughput with the lowest error rates (142, 143).

The use of culture independent methods for identifying and quantifying microbial species is increasing due to the speed and reliability of these methods for identifying unusual and difficult to culture organisms, as well as those in low abundance (144). However, these DNA-based methods have their own drawbacks and areas of potential bias, which are particularly pertinent to the study of CF microbial community diversity. These have to be addressed in order to be confident in the results.

Unlike culture based methods, DNA based analyses are unable to differentiate between DNA from living organisms and extracellular DNA or DNA from dead or damaged cells (145). This has massive implications, both for drawing ecological conclusions about the microbial community within an environment (146), and when considering clinical infections (147). In order to address this, samples can be treated with propidium monoazide (PMA), a

membrane impermeable dye. PMA covalently binds upon exposure to light to extracellular DNA or DNA within cells that have damaged cell membranes (146). Once the PMA binds to free DNA, the DNA is rendered insoluble and as a result it is removed during the extraction procedure, allowing analysis of only the viable community (145).

1.11.4 Propidium monoazide

Discrimination between live and dead cells is important in order to analyse microbial communities and understand how perturbations can effect or change the community dynamics within the CF lung. Membrane integrity as inferred by DNA binding dyes is considered the most practical method for this and is widely used in microscopy and flow cytometry (145).

In some early applications ethidium monoazide (EMA) was used in live-dead cell assays as it intercalates with DNA, covalently bonding to the DNA when exposed to visible light (148). EMA can penetrate cells with compromised cell walls or membranes, allowing binding to DNA within compromised or non-viable cells (145). Nogva *et al* (2003) utilised the photoactivated DNA binding properties of EMA to inhibit DNA from dead bacteria from amplification by PCR (148). This method was found to be promising however, it had a major disadvantage as EMA was found to readily penetrate the membrane of some viable cells, for example *Escherichia coli* (149).

The lack of selectivity observed with the use of EMA lead Nocker *et al* (2006) to test PMA as an alternative (145). PMA was derived from propidium iodide (PI), with the addition of an azide group to allow DNA cross-linking when exposed to light (145). The two positive charges on PMA, compared to the single charge on EMA, are thought to contribute to the reduction in permeability of PMA across live cell membranes resulting in significantly higher selectivity (145).

PMA cross-linking, shown schematically in Figure 1.5, is carried out by exposing microbial cells of interest to PMA in dark conditions for a period of 5-15 min Evidence suggests that

incubation time does not affect DNA yield, however it was noted by Nocker *et al* (2006) that after 15min there was a loss in *E. coli* DNA when PMA was in high concentration (145). This study recommended 5 minutes incubation in the dark prior to light exposure, allowing PMA to covalently bind to DNA. It also showed that 120 seconds of light exposure ensured efficient binding of PMA. Any unbound PMA after the incubation period was also found to be inactivated during this time (145).

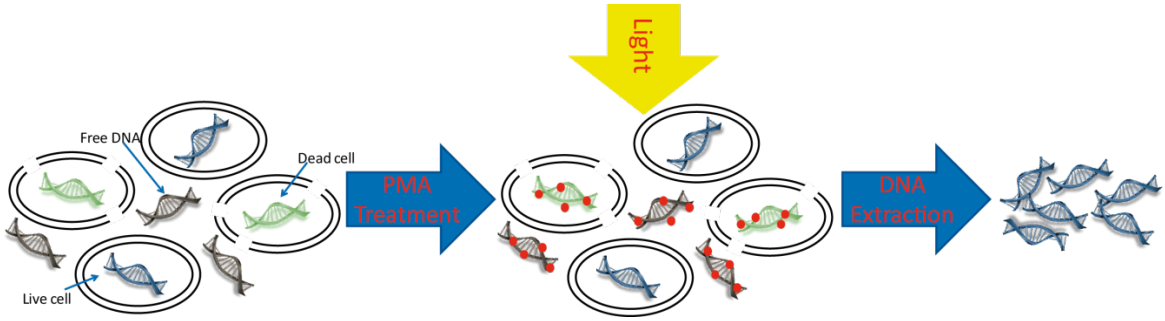


Figure 1.5 Representation of the selective detection of live cells using propidium monoazide.

PMA is added to the sample of interest and incubated in the dark. After incubation samples are exposed to light, resulting in PMA covalently binding to extracellular DNA and DNA from dead or damaged cells.

To investigate the effectiveness of PMA treatment Nocker *et al,* carried out a range of studies on microbial communities from various environmental systems including wastewater and marine sediment (146) and canal and seawater (150). These studies established the effectiveness of PMA treatment for suppressing the DNA signals from extracellular DNA and DNA from dead or damaged cells.

Since the publication of Nocker's work on PMA, this approach has been used in wide range of studies to investigate viable; bacterial (151), fungal (152) or viral (153) communities. Of particular importance is that, PMA has been widely investigated by Rogers *et al,* for use in the investigation of the microbial community in CF sputum. Rogers *et al* (2008) demonstrated that PMA effectively prevented unwanted DNA from being utilized as a template during PCR (144). Investigation into the effects of PMA treatment on community composition revealed that PMA treatment resulted in an increase in community evenness, by increasing community diversity (154). This study also suggested that failure to remove extracellular DNA and DNA from dead and damaged cells could lead to skewed results and misinterpretation of the microbial community composition (154).

1.12 Microbial ecology and CF lung infections

As described above, chronic respiratory infections have been the focus of CF lung disease research for many years. Previous studies using both culture and molecular methodologies have established that a complex and dynamic ecosystem exists within the lung (132, 154, 155). Early investigations into the whole bacterial community within the CF lung were dominated by Rogers *et al*. They initially used T-RFLP and 16S rRNA gene sequencing to identify the composition and diversity of the bacterial community in 2003 (43). Papers published by the same group, in 2004 (156) and 2005 (131) confirm these results. The identification of species not previously identified as part of the CF lung community lead to the

investigation of areas of potential contamination and in 2006, Rogers *et al*, revealed evidence that members of the oral microbiota isolated from sputum were not a result of contamination but were in fact members of the lower respiratory community (157).

Studies by Rogers *et al,* provided a foundation for a large number of other studies, further improving our understanding of the community of organisms associated with CF lung disease. Even so, there is still work to be done in order to translate the results of these studies into improved patient outcomes. Investigations into community structure, using recognised ecological measurements (described in detail in Chapter 2.8), are being carried out in order to allow us to predict how the community will react to perturbations within the system. For example, work carried out by van der Gast *et al* (2011) showed a significant relationship between species richness and lung function, suggesting that patients with more diverse bacterial communities have better disease outcomes (4). This finding was supported by results from Zhao *et* al (2012) in a longitudinal study of CF patients over the course of a year (140).

Although it is clear that a diverse microbial community is present within the CF lung, we are yet to understand the full complexity of even the bacterial community alone. Using a combination of ecological measures, designed to assess species diversity and community composition, and mixed effects modelling, this study aimed to concentrate sequencing efforts on the bacterial community in an attempt to bridge some of these knowledge gaps.

As molecular methodologies become more widely available, achieving an accurate a picture as possible of this bacterial community is vitally important. Understanding the optimal methodologies for sample handling and laboratory procedures allows us to confidently explore bacterial community dynamics and how these relate to disease state. Studies by Rogers *et al* (2013) have investigated the use of PMA to remove the effect of extracellular DNA and that from dead or damaged cells (154) however, no studies have previously investigated the importance of sample handling for the accurate analysis of CF sputum

bacterial communities. By first understanding where areas of potential bias can be encountered we are able to confidently explore the bacterial community from CF sputum.

Bacterial community dynamics over time, during periods of both stability and exacerbation are vital in order to explore disease progression. Previous studies have looked at limited samples from these different disease states, however, no studies have incorporated multiple samples from all clinical stages. By incorporating multiple samples this study was able establish patterns in the data while accounting for variation both between and within patients.

Finally, to date no large scale studies have been carried out to determine if there are any wider patterns in the bacterial community that may result in disease progression. A large scale, spatial study incorporating sputum samples from clinic across the UK and the USA allows us to stratify the data and identify markers that may indicate improved or worsening disease outcomes.

1.12 Aims

As explained in detail above, current literature has revealed the CF lung to be populated by a complex and diverse bacterial community. The overall aim of this study was to investigate, through the use of high throughput sequencing, the bacterial community diversity and composition within the CF lung and its relationship with patient outcomes. By investigating the complex relationship between the bacterial community and the host related factor this thesis aimed to provide insights that may ultimately influence the treatment of CF lung infections.

In recent years there has been a drive towards the use of culture independent technologies for the investigation of bacterial community within the CF lung disease however, few studies have fully assess the impacts of sampling practices on detection of bacterial species. As such, the initial research chapters (Chapter 3, Chapter 4 and Chapter 5) will focus on

methodologies designed to limit bias in the investigation of the bacterial community within the lung. Chapter 3 aims to assess the impact of extracellular DNA and DNA from dead or damaged cell in the analysis of CF bacterial communities, while Chapter 4 focuses on the impact of the initial treatment and storage of sputum samples prior to high throughput sequencing. Chapter 5 aims to address the impact of multiple freeze thaw cycles on the bacterial community in order to minimise bias in the study of samples biobanked for use in multiple projects. These studies aimed to provide guidelines on sample handling allowing the most accurate representation possible of the bacterial community within the CF lung to be obtained.

By first establishing a framework of methodologies designed to provide an accurate depiction of the bacterial community within the lung, these methods were then taken forward to the final chapters. These chapters aim was to uncover the relationship between the bacterial community and clinical factors related to patient disease state. Chapter 6 aims to investigate overall relationships between clinical factors related to CF and the bacterial community within the lung, using a large cohort sample set from 11 sites across Europe and North America. By investigating a large and diverse sample cohort this study aimed to uncover significant relationships between clinical factors and the bacterial community.

While investigation of spatial samples allows insight into important clinical factors longitudinal sampling allows investigation into how CF lung disease progresses and ultimately leads to respiratory failure. Chapter 7 aims to understand the changes in bacterial community dynamics over the course of a CF pulmonary exacerbation, allowing novel insight into the full course of exacerbation as well as allowing identification of novel biomarkers of disease.

1.12 References

1. **World Health Organisation.** 2008. The top 10 causes of death, Fact sheet number 310. www.who.int

2. **Monte SV, Paolini NM, Slazak EM, Schentag JJ, Paladino JA.** 2008. Costs of treating lower respiratory tract infections in outpatients. Am J Manag Care **14:**190-196.

3. **LiPuma JJ.** 2010. The Changing Microbial Epidemiology in Cystic Fibrosis. Clin Microbiol Rev **23:**299-323.

4. **van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD.** 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. ISME J **5:**780-791.

5. **Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S.** 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. PNAS **109:**16213-16216.

6. **Stedman K.** 2013. Mechanisms for RNA Capture by ssDNA Viruses: Grand Theft RNA. J Mol Evol **76:**359-364.

7. **Hawksworth DL.** 2001. Mushrooms: The Extent of the Unexplored Potential. Int J Med Mushrooms **3:**5.

8. **Ehrlich HL.** 1998. Geomicrobiology: its significance for geology. Earth Sci Rev **45:**45-60.

9. **Hooke R.** 1665. Micrographia: or, Some physiological descriptions of minute bodies made by magnifying glasses: with observations and inquiries thereupon. Printed by J. Martyn and J. Allestry, London.

10. **Gest H.** 2004. The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. Notes and Records of the Royal Society of London **58:**187-201.

11. **Kaufmann SHE, Schaible UE.** 2005. 100th anniversary of Robert Koch's Nobel Prize for the discovery of the tubercle bacillus. Trends Microbiol **13:**469-475.

12. **Koch B.** 1905. Über den derzeitigen Stand der Tuberkulosebekämpfung. Nordiskt Medicinskt Arkiv **38:**1-13.

13. **Fredricks DN, Relman DA.** 1996. Sequence-Based Identification of Microbial Pathogens: a Reconsideration of Koch's Postulates. Clin Microbiol Rev **9:**18-33.

14. **Watson JD, Crick FH.** 1953. The structure of DNA. Cold Spring Harbor symposia on quantitative biology **18:**123-131.

15. **Zuckerkandl E, Pauling L.** 1965. Molecules as documents of evolutionary history. J theor biol **8:**357-366.

16. **Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, Pechman K, Lewis BJ, Stahl D.** 1975. Conservation of primary structure in 16S ribosomal RNA. Nature **254:**83-86.

17. **Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindal BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk H, Eisen JA.** 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature **462**: 1056:160

18. **Stewart EJ.** 2012. Growing Unculturable Bacteria. J Bacteriol **16**: 4151:4160

19. **Janda JM, Abbott SL.** 2007. 16S rRNA Gene Sequencing for bacterial Identification in the Diagnositic Laboratory: Pliuses, Perils and Pitfalls. J Clin Microbiol **45**(9):2761-2764

20. **Weisburg WG, Barns SM, Pelletier DA, Lane DJ.** 1991. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol **173:**697-703.

21. **Coenye T, Vandamme P.** 2003. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiol Lett **228:**45-49.

22. **Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH.** 2008. Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. Evol bioinf online **4:**193-201.

23. **Martin KJ, Rygiewicz PT.** 2005. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. BMC Microbiol **5**.28-39

24. **Mallia P, Footitt J, Sotero R, Jepson A, Contoli M, Trujillo-Torralbo MB, Kebadze T, Aniscenko J, Oleszkiewicz G, Gray K, Message SD, Ito K, Barnes PJ, Adcock IM, Papi A, Stanciu LA, Elkin SL, Kon OM, Johnson M, Johnston SL.** 2012. Rhinovirus Infection Induces Degradation of Antimicrobial Peptides and Secondary Bacterial Infection in Chronic Obstructive Pulmonary Disease. Am J Resp Crit Care **186:**1117-1124.

25. **Rosario K, Breitbart M.** 2011. Exploring the viral world through metagenomics. Curr Opin Virol **1:**289-297.

26. **Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, Green JL, Green LE, Killham K, Lennon JJ, Osborn AM, Solan M, van der Gast CJ, Young JPW.** 2007. Essay - The role of ecological theory in microbial ecology. Nat Rev Microbiol **5:**384-392.

27. **Magurran AE.** 2004. Measuring biological diversity. Blackwell Pub, Malden, Ma.

28. **Mayr E**. 1957. The Species Problem. American Association for the advancement of science, Washington DC. 1-22

29. **Green LE, Medley GF**. 2002. Mathematical modelling of the foot and mouth disease epidemic of 2001: strengths and weaknesses. Res Vet Sci **73**(3):201-205

30. **Wu Y, Zhu W, Lu X**. 2013. Identifying key parameters in a novel multistep bio-ecological wastewater treatment process for rural areas. Ecol Eng **61**:166-173

31. **Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA.** 2012. The Application of Ecological Theory Towards an Understanding of the Human Microbiome. Science **336**:125-262

32. **Moyle PB, Light T.** 1996. Biological Invasions of Fresh Water: Empirical Rules and Assembly Theory. Biol Conserv **78**:149-161

33. **Wilson DS**.1992. Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. Ecol **73**:1984-2000

34. **Chan JZ, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ.** 2012. Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol **12:**302-313

35. **Konstantinidis KT, Ramette A, Tiedje JM.** 2006. The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci **361(1475):**1929-1940.

36. **Janda JM, Abbott SL.** 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol **45:**2761-2764.

37. **Stackebrandt EE, J.** 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today **33:**152–155.

38. **Hong S, Bunge J, Leslin C, Jeon S, Epstein SS.** 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J **3:**1365-1373.

39. **Gans J, Wolinsky M, Dunbar J.** 2005. Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. Science **309:**1387-1390.

40. **Shendure J, Aiden EL.** 2012. The expanding scope of DNA sequencing. Nat Biotech **30:**1084-1094.

41. **Schütte UE, Abdo Z, Bent S, Shyu C, Williams C, Pierson J, Forney L.** 2008. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Appl Microbiol Biotechnol **80:**365-380.

42. **Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ, Bruce KD.** 2003. Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. J Clin Microbiol **41:**3548-3558.

43. **Dunbar J, Ticknor LO, Kuske CR.** 2000. Assessment of Microbial Diversity in Four Southwestern United States Soils by 16S rRNA Gene Terminal Restriction Fragment Analysis. Appl Environ Microbiol **66:**2943-2950.

44. **Engebretson JJ, Moyer CL.** 2003. Fidelity of Select Restriction Endonucleases in Determining Microbial Diversity by Terminal-Restriction Fragment Length Polymorphism. Appl Environ Microbiol **69:**4823-4829.

45. **Rogers GB, Bruce KD.** 2010. Next-Generation Sequencing in the Analysis of Human Microbiota Essential Considerations for Clinical Application. Mol Diagn Ther **14:**343-350.

46. **Schuster SC.** 2008. Next-generation sequencing transforms today's biology. Nat methods **5:**16-18.

47. **Rothberg JM, Leamon JH.** 2008. The development and impact of 454 sequencing. Nat Biotech **26:**1117-1124.

48. **Degnan PH, Ochman H.** 2012. Illumina-based analysis of microbial community diversity. ISME J **6:**183-194.

49. **Sleator RD.** 2010. The human superorganism – Of microbes and men. Med Hypotheses **74:**214-215.

50. **Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI.** 2007. The Human Microbiome Project. Nature **449:**804-810.

51. **Cho I, Blaser MJ.** 2012. APPLICATIONS OF NEXT-GENERATION SEQUENCING The human microbiome: at the interface of health and disease. Nat Rev Genet **13:**260-270.

52. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the aquasition and structure of the initial microbiota across multiple body habitats in newborns. PNAS 107:11971-11975

53. **Brailsford SR**. 2005. The microflora of the erupting first permenant molar. Caries Res **39**:78-84

54. **Gao Z, Tseng CH, Strober EB, Pei Z, Blaser MJ**. 2008. Substamntial alterations of the cutaneous bacterial biota in psoriatic leasions. PLoS ONE **3**:e2719

55. **Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI**. 2006. An obesity associated gut microbiome with increased capacity for energy harvest. Nature. **444**:1027-1031

56. **Blaser MJ, Chen Y, Reibman J.** 2008. Does *Helicobacter pylorli* protect against asthma and allergy? Gut **57**:561-567

57. **Garrett WS, Gallini CA, Yatsunenko T, Michaud M, DuBois A, Delaney ML, Punit S, karlsson M, Bry L, Glickman JN, Gordon JI, Onderdonk AB, Glimcher LH**. 2010. *Enterobacteriaceae* act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. Cell Host Microbe **8**:292-300

58. **Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, Feldstein AE, Britt EB, Fu X, Chung Y, Wu Y, Schauer P, Smith JD, Allayee H, Tang WHW, DiDonto JA, Lusis AJ, Hazen SL.** 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature **22**:57-63

59. **The Human Microbiome Project**. 2012. Structure, function and diversity of the healthy human microbiome. Nature **486**:207-214

60. **Wilson M, Wilson M.** 2004. The respiratory system and its indigenous microbiota. Microbial Inhabitants of Humans. Cambridge University Press.

61. **Mims CA, Nash A, Stephen J.** 2001. Mims' Pathogenesis of infectious disease, 5th ed. Academic Press, San Diego.

62. **Owen MK, Ensor DS, Sparks LE.** 1992. Airborne particle sizes and sources found in indoor air. Atmos Environ, Part A, General Topics **26:**2149-2162.

63. **Knox RB, Suphioglu C, Taylor P, Desai R, Watson HC, Peng JL, Bursill LA.** 1997. Major grass pollen allergen Lol p 1 binds to diesel exhaust particles: implications for asthma and air pollution. Clin Exp Allergy **27:**246-251.

64. **Brodie EL, DeSantis TZ, Parker JPM, Zubietta IX, Piceno YM, Andersen GL.** 2007. Urban aerosols harbor diverse and dynamic bacterial populations. PNAS USA **104:**299-304.

65. **DiGiorgio C, Krempff A, Guiraud H, Binder P, Tiret C, Dumenil G.** 1996. Atmospheric pollution by airborne microorganisms in the city of Marseilles. Atmos Environ **30:**155-160.

66. **Rautemaa R, Nordberg A, Wuolijoki-Saaristoe K, Meurman JH.** 2006. Bacterial aerosols in dental practice - a potential hospital infection problem? J Hosp Infect **64:**76-81.

67. **Boucher RC.** 2007. Evidence for airway surface dehydration as the initiating event in CF airway disease. J Intern Med **261:**5-16.

68. **Nicod LP.** 2005. Lung defences: an overview. Eur Resp Rev **14:**45-50.

69. **Singh PK, Jia HP, Wiles K, Hesselberth J, Liu L, Conway B-AD, Greenberg EP, Valore EV, Welsh MJ, Ganz T, Tack BF, McCray PB.** 1998. Production of β-defensins by human airway epithelia. PNAS **95:**14961-14966.

70. **Hunninghake GW, Gadek JE, Kawanami O, Ferrans VJ, Crystal RG.** 1979. Inflammatory and immune processes in the human lung in health and disease: evaluation by bronchoalveolar lavage. Am J Pathol **97:**149-206.

71. **Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG.** 2011. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med **184:**957-963.

72. **Lees AW, McNaught W.** 1959. Bacteriology of lower-respiratory-tract secretions, sputum, and upper-respiratory-tract secretions in "normals" and chronic bronchitics. Lancet **2:**1112-1115.

73. **Cabello H, Torres A, Celis R, ElEbiary M, delaBellacasa JP, Xaubet A, Gonzalez J, Agusti C, Soler N.** 1997. Bacterial colonization of distal airways in healthy subjects and chronic lung disease: A bronchoscopic study. Eur Respir J **10:**1137-1144.

74. **Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, Martinez FJ, Huffnagle GB.** 2011. Analysis of the Lung Microbiome in the "Healthy" Smoker and in COPD. PLOS One **6**.

75. **Huang YJ, Nelson CE, Brodie EL, Desantis TZ, Baek MS, Liu J, Woyke T, Allgaier M, Bristow J, Wiener-Kronish JP, Sutherland ER, King TS, Icitovic N, Martin RJ, Calhoun WJ, Castro M, Denlinger LC, Dimango E, Kraft M, Peters SP, Wasserman SI, Wechsler ME, Boushey HA, Lynch SV.** 2011. Airway microbiota

and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. J Allergy Clin Immunol **127:**372-381 e371-373.

76. **Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF, Cookson WOC.** 2010. Disordered Microbial Communities in Asthmatic Airways. PLOS One **5:**e8578.

77. **Woodhead M, Blasi F, Ewig S, Garau J, Huchon G, Ieven M, Ortqvist A, Schaberg T, Torres A, van der Heijden G, Read R, Verheij TJM, Joint Taskforce of the European Respiratory S, European Society for Clinical M, Infectious D.** 2011. Guidelines for the management of adult lower respiratory tract infections - Full version. Clin Microbiol Infec **17:**E1-E59.

78. **Fendrick A, Monto AS, Nightengale B, Sarnes M.** 2003. THe economic burden of non–influenza-related viral respiratory tract infection in the united states. Arch Intern Med **163:**487-494.

79. **NICE Short Clinical Guidelines Technical Team and members of the GDG.** 2008. Respiratory tract infections – antibiotic prescribing. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. London: NICE (UK)

80. **Busch R.** 1990. On the history of cystic fibrosis. Acta Univ Carol Med **36:**13-15.

81. **Rohlfs EM, Zhou ZQ, Heim RA, Nagan N, Rosenblum LS, Flynn K, Scholl T, Akmaev VR, Sirko-Osadsa DA, Allitto BA, Sugarman EA.** 2011. Cystic Fibrosis Carrier Testing in an Ethnically Diverse US Population. Clin Chem **57:**841-848.

82. **Cystic Fibrosis Foundation.** 2012. http://www.cff.org/AboutCF/

83. **UK CF Registry.** 2010. Annual Data Report 2010.

84. **Bilton D.** 2008. Cystic fibrosis. Medicine **36:**273-278.

85. **Lyczak JB, Cannon CL, Pier GB.** 2002. Lung infections associated with cystic fibrosis. Clin Microbiol Rev **15:**194-222.

86. **Govan JRW, Deretic V.** 1996. Microbial pathogenesis in cystic fibrosis: Mucoid Pseudomonas aeruginosa and Burkholderia cepacia. Microbiol Rev **60:**539-574.

87. **Boucher RC.** 2004. New concepts of the pathogenesis of cystic fibrosis lung disease. Eur Respir J **23:**146-158.

88. **Flume PA, Mogayzel PJ, Robinson KA, Rosenblatt RL, Quittell L, Marshall BC, Pulm CPG.** 2010. Cystic Fibrosis Pulmonary Guidelines Pulmonary Complications: Hemoptysis and Pneumothorax. Am J Resp Crit Care **182:**298-306.

89. **Andersen DH.** 1938. Cystic fibrosis of the pancreas and its relation to celiac disease: A clinical and pathologic study. Am J Dis Child **56:**344-399.

90. **Andersen DH, Hodges RG.** 1946. Celiac syndrome; genetics of cystic fibrosis of the pancreas, with a consideration of etiology. Am J Dis Child **72:**62-80.

91. **Di Sant'Agnese PA, Darling RC, Perera GA, Shea E.** 1953. Abnormal electrolyte composition of sweat in cystic fibrosis of the pancreas; clinical significance and relationship to the disease. Pediatrics **12:**549-563.

92. **Davis PB.** 2006. Cystic fibrosis since 1938. Am J Respir Crit Care Med **173:**475-482.

93. **Gibson LE, Cooke RE.** 1959. A test for concentration of elecyrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. Pediatrics **23:**545-549.

94. **Knowles MR, Stutts MJ, Spock A, Fischer N, Gatzy JT, Boucher RC.** 1983. Abnormal Ion Permeation through Cystic-Fibrosis Respiratory Epithelium. Science **221:**1067-1070.

95. **Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, et al.** 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science **245:**1059-1065.

96. **Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al.** 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science **245:**1066-1073.

97. **Kerem BS, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC.** 1989. Identification of the Cystic-Fibrosis Gene: Genetic Analysis. Science **245:**1073-1080.

98. **Zobell JT, Young DC, Waters CD, Stockmann C, Ampofo K, Sherwin CMT, Spigarelli MG.** 2012. Optimization of anti-pseudomonal antibiotics for cystic fibrosis pulmonary exacerbations: I. aztreonam and carbapenems. Pediatr Pulm **47:**1147-1158.

99. **Ahmad A, Ahmed A, Patrizio P.** 2013. Cystic fibrosis and fertility. Curr Opin Obstet Gynecol **25:**167-172

100. **Weintraub A, Blau H, Mussaffi H, Picard E, Bentur L, Kerem E, Stankiewicz H, Wilschanski M.** 2009. Exocrine Pancreatic Function Testing in Patients With Cystic Fibrosis and Pancreatic Sufficiency: A Correlation Study. J Pediatr Gastr Nutr **48:**306-310.

101. **Cheng K, Ashby D, Smyth RL.** 2012. Ursodeoxycholic acid for cystic fibrosis-related liver disease. Cochrane Database Syst Rev **10**.

102. **Zirbes J, Milla CE.** 2009. Cystic fibrosis related diabetes. Paediatr Respir Rev **10:**118-123.

103. The Clinical and Functional Translation of CFTR (CFTR2). @ Copyright 2011 US CF Foundation, Johns Hopkins University, The Hospital for Sick Children. www.CFTR2.org.

104. **Zielenski J, Tsui LC.** 1995. Cystic fibrosis: Genotypic and phenotypic variations. Annu Rev Genet **29:**777-807.

105. **Geborek A, Hjelte L**. 2011. Association between genotype and Pulmonary phenotype in cystic fibrosis patients with severe mutations. Journal of Cystic Fibrosis. **10**(3):187-192

106. **McKone EF, Goss CH, Aitken ML.** 2006. CFTR genotype as a predictor of prognosis in cystic fibrosis. CHEST **130:**1441-1447.

107. **Southern KW.** 2007. Cystic Fibrosis and Formes Frustes of CFTR-Related Disease. Respiration **74:**241-251.

108. **Lukacs GL, Verkman AS**. 2012. CFTR: folding, missfolding and correcting the **ΔF508** conformational defect. Trends Mol Med **18**(2):81-91

109. **Ratjen F.** 2007. New pulmonary therapies for cystic fibrosis. Curr Opin Pulm Med **13:**541-546.

110. **Smith JJ, Travis SM, Greenberg EP, Welsh MJ.** 1996. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. Cell **85:**229-236.

111. **Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW, Boucher RC.** 1998. Evidence for Periciliary Liquid Layer Depletion, Not Abnormal Ion Composition, in the Pathogenesis of Cystic Fibrosis Airways Disease. Cell **95:**1005-1015.

112. **Mall M, Grubb BR, Harkema JR, O'Neal WK, Boucher RC.** 2004. Increased airway epithelial Na+ absorption produces cystic fibrosis-like lung disease in mice. Nat Med **10:**487-493.

113. **Bonfield TL, Konstan MW, Burfeind P, Panuska JR, Hilliard JB, Berger M.** 1995. Normal bronchial epithelial cells constitutively produce the anti-inflammatory cytokine interleukin-10, which is downregulated in cystic fibrosis. Am J Resp Cell Mole Biol **13:**257-261.

114. **Cohen-Cymberknoh M, Kerem E, Ferkol T, Elizur A.** 2013. Airway inflammation in cystic fibrosis: molecular mechanisms and clinical implications. Thorax 68:1157-1162

115. **Balough K, McCubbin M, Weinberger M, Smits W, Ahrens R, Fick R.** 1995. The relationship between infection and inflammation in the early stages of lung disease from cystic fibrosis. Pediatr Pulmonol **20:**63-70.

116. **Ranganathan SC, Parsons F, Gangell C, Brennan S, Stick SM, Sly PD and on bahalf of AREST CF.** 2011. Evolution of pulmonary inflammation and nutritional status in infants and young children with cystic fibrosis. Thorax **66:**408-413.

117. **Khan TZ, Wagener JS, Bost T, Martinez J, Accurso FJ, Riches DW.** 1995. Early pulmonary inflammation in infants with cystic fibrosis. Am J Respir Crit Care Med **151:**1075-1082.

118. **Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM.** 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. ISME J **5:**20-29.

119. **Harrison F.** 2007. Microbial ecology of the cystic fibrosis lung. Microbiol **153:**917-923.

120. **Rosenfeld M, Emerson J, Williams-Warren J, Pepe M, Smith A, Montgomery AB, Ramsey B.** 2001. Defining a pulmonary exacerbation in cystic fibrosis. J Pediatr **139:**359-365.

121. **Flume PA, Mogayzel PJ, Robinson KA, Goss CH, Rosenblatt RL, Kuhn RJ, Marshall BC, Pulm CPG.** 2009. Cystic Fibrosis Pulmonary Guidelines Treatment of Pulmonary Exacerbations. Am J Resp Crit Care **180:**802-808.

122. **The UK Cystic Fibrosis Trust Microbiology Laboratory Standards Working Group.** 2010. Laboratory Standards for Processing Microbiological Samples from People with Cystic Fibrosis. Cystic Fibrosis Trust **First edition**.

123. **Belessis Y, Dixon B, Hawkins G, Pereira J, Peat J, MacDonald R, FieldP, Numa A, Morton J, Lui K, Jaffe A.** 2012. Early Cystic Fibrosis Lung Disease Detected by Bronchoalveolar Lavage and Lung Clearance Index. Am J Resp Crit Care **185:**862-873.

124. **Brennan S, Gangell C, Wainwright C, Sly PD.** 2008. Disease surveillance using bronchoalveolar lavage. Paediatr Respir Rev **9:**151-159.

125. **Aaron SD, Kottachchi D, Ferris WJ, Vandemheen KL, St. Denis ML, Plouffe A, Doucette SP, Saginur R, Chan FT, Ramotar K.** 2004. Sputum versus bronchoscopy for diagnosis of Pseudomonas aeruginosa biofilms in cystic fibrosis. Eur Respir J **24:**631-637.

126. **Nicholas B, Djukanovic R.** 2009. Induced sputum: a window to lung pathology. Biochem Soc Trans **37:**868-872.

127. **Rogers GB, Skelton S, Serisier DJ, van der Gast CJ, Bruce KD.** 2010. Determining cystic fibrosis-affected lung microbiology: comparison of spontaneous and serially induced sputum samples by use of terminal restriction fragment length polymorphism profiling. J Clin Microbiol **48:**78-86.

128. **Aitken ML, Greene KE, Tonelli MR, Burns JL, Emerson JC, Goss CH, Gibson RL.** 2003. Analysis of sequential aliquots of hypertonic saline solution-induced sputum from clinically stable patients with cystic fibrosis. Chest **123:**792-799.

129. **van der Gast CJ, Cuthbertson L, Rogers GB, Pope C, Marsh RL, Redding GJ, Bruce KD, Chang AB, Hoffman LR.** 2014. Three Clinically Distinct Chronic Pediatric Airway Infections Share a Common Core Microbiota. Anna Am Thorac Soc.

130. **Henig NR, Tonelli MR, Pier MV, Burns JL, Aitken ML.** 2001. Sputum induction as a research tool for sampling the airways of subjects with cystic fibrosis. Thorax **56:**306-311.

131. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Kehagia V, Jones GR, Bruce KD.** 2005. Bacterial activity in cystic fibrosis lung infections. Respir Res **6:**49.

132. **Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS, Wolfgang MC, Boucher R, Gilpin DF, McDowell A, Elborn JS.** 2008. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am J Resp Crit Care **177:**995-1001.

133. **Borman AM, Palmer MD, Delhaes L, Carrere J, Favennec L, Ranque S, Gangneux JP, Horre R, Bouchara JP.** 2010. Lack of standardization in the procedures for mycological examination of sputum samples from CF patients: a possible cause for variations in the prevalence of filamentous fungi. Med Mycol **48:**S88-S97.

134. **Mogayzel PJ, Naureckas ET, Robinson KA, Brady C, Guill M, Lhiri T, Lubsch L, Matsui J, Oermann CM, Ratjen F, Rosenfeld M, Simon RH, Hazle L, Sabadosa K, Marshal BC, and the Cystic Fibrosis Foundation Pulmonary Clinical Practice Guidelines Committee.** 2014. Cystic Fibrosis Foundation Pulmonary Guielines. Pharmacologic Approaches to Prevention and Eradication of Initial Pseudomonas aeruginosa Infection. Ann Am Thorac Soc **11**(10)**:**1640-1650.

135. **Cystic Fibrosis Foundation Patient Registry.** 2012. 2012 Annual Data Report. Bethesda, Maryland ©2013 Cystic Fibrosis Foundation.

136. **Amann RI, Ludwig W, Schleifer KH.** 1995. Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation. Microbiol Rev **59:**143-169.

137. **Miller DN, Bryant JE, Madsen EL, Ghiorse WC.** 1999. Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples. Appl Environ Microbiol **65:**4715-4724.

138. **Pitcher DG, Saunders NA, Owen RJ.** 1989. Rapid Extraction of Bacterial Genomic DNA with Guanidium Thiocyanate. Lett Appl Microbiol **8:**151-156.

139. **Nagano Y, Elborn JS, Miller BC, Walker JM, Goldsmith CE, Rendall J, Moore JE.** 2010. Comparison of techniques to examine the diversity of fungi in adult patients with cystic fibrosis. Med Mycol **48:**166-176.

140. **Zhao JC, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ.** 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. PNAS USA **109:**5809-5814.

141. **Maughan H, Wang PW, Caballero JD, Fung P, Gong YC, Donaldson SL, Yuan LJ, Keshavjee S, Zhang Y, Yau YCW, Waters VJ, Tullis DE, Hwang DM, Guttman DS.** 2012. Analysis of the Cystic Fibrosis Lung Microbiota via Serial Illumina Sequencing of Bacterial 16S rRNA Hypervariable Regions. PLOS One **7**(10):e45791.

142. **Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ**. 2012. Performance comparison of benchtop high-troughput sequencing platforms. Nat Biotech **30**(5):434-439

143. **Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D**. 2013. Updating benchtop sequencing performance comparison. Nat Biotech **31**(4):294-296

144. **Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD.** 2008. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diagn Microbiol Infect Dis **62:**133-141.

145. **Nocker A, Cheung CY, Camper AK.** 2006. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. J Microbiol Methods **67:**310-320.

146. **Nocker A, Sossa-Fernandez P, Burr MD, Camper AK.** 2007. Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol **73:**5111-5117.

147. **Nocker A, Sossa KE, Camper AK.** 2007. Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR. J Microbiol Methods **70:**252-260.

148. **Nogva HK, Dromtorp SM, Nissen H, Rudi K.** 2003. Ethidium monoazide for DNA-based differentiation of viable and dead bacteria by 5 '-nuclease PCR. Biotechniques **34:**804-813.

149. **Nocker A, Camper AK.** 2006. Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. Appl Environ Microbiol **72:**1997-2004.

150. **Nocker A, Richter-Heitmann T, Montijn R, Schuren F, Kort R.** 2010. Discrimination between live and dead cellsin bacterial communities from environmental water samples analyzed by 454 pyrosequencing. Int Microbiol **13:**59-65.

151. **Takahashi Y, Yoshida A, Nagayoshi M, Kitamura C, Nishihara T, Awano S, Ansai T.** 2011. Enumeration of viable Enterococcus faecalis, a predominant apical periodontitis pathogen, using propidium monoazide and quantitative real-time polymerase chain reaction. Microbiol Immunol **55:**889-892.

152. **Vesper S, McKinstry C, Hartmann C, Neace M, Yoder S, Vesper A.** 2008. Quantifying fungal viability in air and water samples using quantitative PCR after treatment with propidium monoazide (PMA). J Microbiol Meth **72:**180-184.

153. **Parshionikar S, Laseke I, Fout GS.** 2010. Use of Propidium Monoazide in Reverse Transcriptase PCR To Distinguish between Infectious and Noninfectious Enteric Viruses in Water Samples. Appl Environ Microbiol **76:**4318-4326.

154. **Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, Lilley AK, Oliver A, Carroll MP, Bruce KD, van der Gast CJ.** 2013. Reducing bias in bacterial community analysis of lower respiratory infections. ISME J **7:**697-706.

155. **Stressmann F, Rogers G, Marsh P, Lilley A, Daniels T, Carroll M, Hoffman L, Jones G, Allen C, Patel N, Forbes B, Tuck A, Bruce K.** 2011. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? J Cyst Fibros **10:**357 - 365.

156. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Bruce KD.** 2004. Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. J Clin Microbiol **42:**5176-5183.

157. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, Connett GJ, Bruce KD.** 2006. Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. J Clin Microbiol **44:**2601-2604.

Chapter 2: Materials and methods

Materials and methods

"Truth has nothing to do with the conclusion, and everything to do with the methodology."

Stefan Molyneux

2.1 Bacterial strains

Bacterial type strains and clinical isolates were used in this study as controls; details of individual bacterial strains used are shown in Table 2.1. Bacterial strains used were chosen due to their clinical relevance in CF and other respiratory infections.

Strains were grown by adding one bead from a frozen stock of the strain to 5ml of Luria Broth (LB) (Oxoid, Hampshire, UK) for 24 hours at 37°C with shaking at 100rpm. To confirm strain purity 100ul of LB (Oxoid, Hampshire, UK) was streaked onto LB agar (Oxoid, Hampshire, UK) and grown for 24 hours at 37°C. A single pure colony was then selected and grown in LB (Oxoid, Hampshire, UK) for 24 hours at 37°C with shaking at 100rpm. DNA extractions were performed as described below.

Mock communities were made for use as sequencing controls in order to account for sequencing bias. PCRs were performed, as indicated in Chapter 2.4.2, on DNA from pure strains of *Pseudomonas aeruginosa, Achromobacter xylosoxidans, Stenotrophomonas maltophilia, Staphylococcus aureus* and *Burkholderia cenocepacia*. Known concentrations (100 ng/µl, calculated using the NanoDrop 8000 (Thermo Fisher Scientific, Loughborough, UK)) of amplified DNA was then combined in sterile microcentrifuge tubes and cleaned-up as indicated in Chapter 2.5. Pure culture bacterial and mock communities controls were included in sequencing runs.

Table 2.1 List of bacterial control strains and sources.

Relevant bacterial controls, isolated from CF sputum were used in all experiments to confirm expected results.

Bacterial controls	Isolated from	Identity confirmed by	Source
Pseudomonas aeruginosa	Adult CF sputum	16S sequencing	Damien Rivett, KCL
Achromobacter xylosoxidans	Adult CF sputum	16S sequencing	Damien Rivett, KCL
Stenotrophomonas maltophilia	Adult CF sputum	16S sequencing	Damien Rivett, KCL
Staphylococcus aureus	Adult CF sputum	16S sequencing	Damien Rivett, KCL
Burkholderia cenocepacia (J2315)	Adult CF sputum	recA PCR/ pulse field electrophoresis/ sequencing	Prof. John Govan, University of Edinburgh
Burkholderia maltophilia (J2337)	Adult CF sputum	recA PCR/ pulse field electrophoresis/ sequencing	Prof. John Govan, University of Edinburgh

2.2 Clinical samples

All sputum samples were collected under full ethical approval, details of ethics associated with each chapter will be indicated in the condensed methods present in each chapter. An over view of ethics details are presented in Table 2.2. Numbers of samples and collection location varied between chapters, patient details of the exclusion criteria are included in the methods of individual chapters. Patient consent was gained before the collection of all samples.

Spontaneous expectorated sputum from patients with cystic fibrosis was used as the sampling method for all studies. Sputum samples were collected in sterile containers and stored at -80°C prior to further processing. In all chapters, with the exception of Chapter 7, sputum was collected during routine clinical visits to local CF centres. If required, the patient provided separate sputum samples for routine microbiology. Samples collected in Chapter 7 were collected by patients in their home and collected by motorcycle courier and returned to Southampton general hospital for storage at -80°C.

Health professionals within the hospital collected associated patient metadata at the time of sample collection. All data and samples were anonymised in accordance with ethical requirements.

Table 2.2 List of ethics reference numbers.

Ethics was obtained for the collection of sputum samples from CF patients for the study of the microbial community within the lung.

Hospital	**Ethics Review board**	**Reference number**
Southampton General Hospital	Southampton and South West Hampshire Research Ethics Committee	06/Q1704/26
Belfast City Hospital	Office for Research Ethics Northern Ireland	06/NIR01/11
Belfast City Hospital; Vectura, samples from the UK, Eire and Poland	1. Multi-Centre Research Ethics Committee for Wales (UK) gave approval on 18th March 2008 2. Irish Ethics Committee (St Vincent's Healthcare Group Ltd Ethics) approved the study on 11th November 2008 3. Polish Ethics Committee (Bioethics Committee of the Medical University, Lodz) final approval gained on the 14th July 2009 (previously stated as the 21st April 2009)	VR496/005
Geisel School of Medicine at Dartmouth	Geisel School of Medicine and Dartmouth College Institutional Review Board	CPHS # 23809
Maine Medical Centre	Maine Medical Center Institutional Review Board	IRB # 4170
Seattle Children's Hospital	Seattle Children's Hospital Institutional Review Board	IRB #12811
Vermont Medical centre	University of Vermont Institutional Review Board	M13-160
Massachusetts General Hospital	Massachusetts General Hospital institutional review board	2011P000620

2.3 Sputum sample processing

2.3.1 Sputum wash

Based on Rogers *et al* (2003) (1), a sterile scalpel was used to transfer approximately 500 µg of frozen sputum into a sterile 15 ml centrifuge tube. Sputum samples were washed three times with sterile 0.8% w/v phosphate-buffered saline (PBS) (Oxoid, Hampshire, UK) to remove saliva. Washing was carried by vigorously shaking samples by hand, with 5 ml of PBS then centrifuging at 4109 x *g* for 5 min. The supernatant was removed and the process was repeated twice more with 500µl PBS. Washed sputum was re-suspended in 500 µl of PBS.

2.3.2 Propidium Monoazide treatment

PMA treatment was performed as previously described by Nocker *et al* (2007) and Rogers *et al* (2008) in order to exclude free DNA and DNA from non-viable cells from further analysis (2, 3).

PMA (Biotium, Hayward, CA) was diluted with 20% dimethyl sulphoxide (DMSO) (Sigma-Aldrich, Dorset, UK) to create a 20 mM PMA stock solution, which was stored in the dark at -20°C. Sputum suspended in 500 µl of PBS was added to black 1.5 ml micro-centrifuge tubes (TreffLab, Degersheim, Switzerland). In a darkened room, 1.25 µl PMA stock solution was added to 500 µl of sample (in black micro-centrifuge tubes) to give a final concentration of 50 µM. Samples were vigorously mixed (10-15 sec) prior to incubation at room temperature in the dark for 15 min, vortexing every 2-3 mins to ensure even mixing. After incubation, contents were transferred to clear micro-centrifuge tubes and placed in an LED light box (see Figure 2.1) for 15 min vortexing every 2-3 mins, in order ensure all PMA molecules were covalently bound to free DNA molecules.

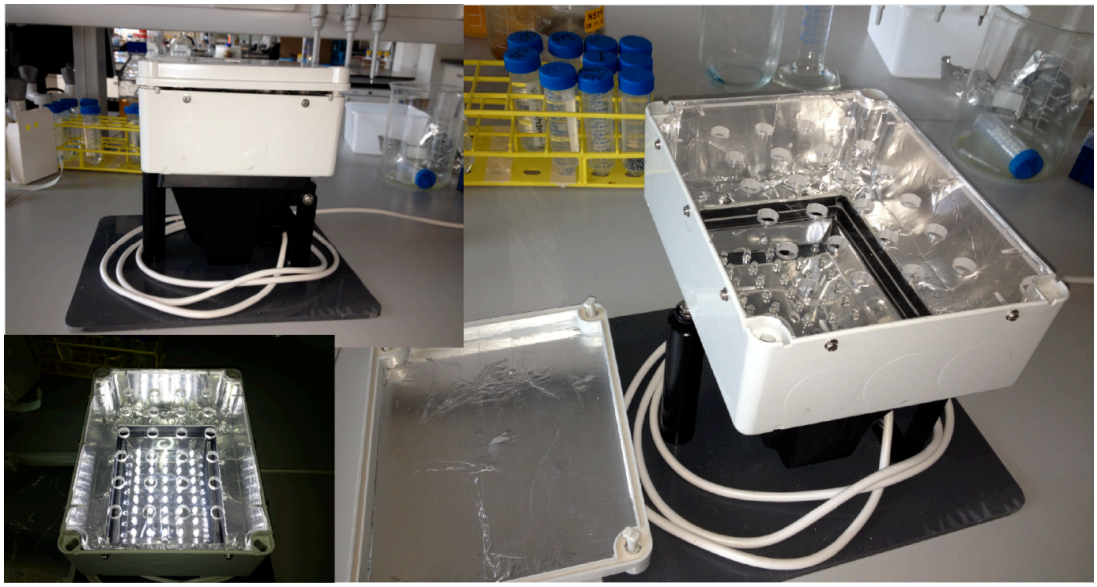


Figure 2.1 Photograph of LED light box.

The light box was constructed by engineering staff at the Centre for Ecology and Hydrology, Wallingford. A 45-LED security floodlight (product number; BB-HL119) was attached to a PVC base. An electoral junction box, lined with reflective tape, was modified to fit over the floodlight. A clear PVC plastic sheet was fitted in the top of the box and holes were drilled to accommodate 24, 1.5 ml microcentrifuge tubes.

2.3.3 DNA extraction

The DNA extraction method used was based on Rogers *et al* (2008) (3). All sputum samples were treated with PMA prior to DNA extraction with the exception of those described in the methods of Chapter 3. Samples were transferred to individual 1.5 ml screw cap microtubes (Sarstedt, Leicester, UK) containing: a single 3 mm diameter tungsten carbide bead (Qiagen, Manchester, UK), 0.2 g of 0.18 mm diameter glass beads (Sigma-Aldrich, Dorset, UK) and 300 µl of guanidinium thiocyanate (5 mol/l), ethylenediaminetetraacetic acid (EDTA)(100mmol/l) -Sarkosyl (0.5% v/v) (GES) (4). Samples were homogenised at 30 Hz in a Mix Mill 300 (Qiagen, Manchester, UK) for 2 x 30 sec, then incubated at 80˚C for 3 min, in order to help dissolve fats and denature proteins not disrupted by physical lysis, samples were then rapidly cooled by chilling at -20˚C for 5 min. Beads and other debris were pelleted by spinning (12,000 X *g* for 5 min) and the supernatant removed and added to micro-centrifuge tubes containing 141 µl of NaCl, to a final concentration of 0.5 mM, and 374 µl polyethylene glycol (PEG) (Fisher scientific, Loughborough, UK), to a final concentration of 15%. DNA was precipitated for 1 hour at 4˚C. Samples were then centrifuged (15,000 X *g* at 4˚C for 10 min) to pellet the DNA. The supernatant was removed and the pellet was resuspended in 500 µl dH_2O, prior to the addition of 500 µl Tris-buffered phenol (pH 8.0) (Invitrogen, Paisley, UK) and vigorous vortexing. The tubes were left to stand for 1min at room temperature then centrifuged for 4min at 15,000 X *g*. The supernatant was added to fresh micro-centrifuge tubes with an equal volume of Tris-buffered phenol (pH 8.0)-chloroform-isoamyl alcohol (1:1) (Sigma-Aldrich, Dorset, UK) and mixed vigorously before centrifugation at 15,000 X *g* for 4 min. The supernatant was removed into fresh microcentrifuge tubes and an equal volume of isopropanol (1:1) (Fisher scientific, Loughborough, UK), 1:10 volume of 10 M ammonium acetate (Ambion, Paisley, UK) and 1 µl of linear polyacrylamide (LPA) (Sigma-Aldrich, Dorset, UK), added and mixed prior to incubation at -20˚C. After 1 hour, samples were centrifuged at 15,000 X *g* for 10 min and the supernatant removed and discarded. The pellet was washed three times with 500 µl 70% ethanol (Fisher scientific, Loughborough, UK). Pellets were air dried, resuspended in 50 µl of

sterile nuclease-free water and stored at -20˚C. For every round of extractions, a sample containing nuclease-free water was extracted in parallel as a negative control.

The presence of DNA was verified by Tris-Borate-EDTA (TBE)(Severn Biotech, Worcester, UK) agarose gel electrophoresis on 1% w/v TBE gels containing 1 µl of 10 mg/ml ethidium bromide (Sigma-Aldrich, Dorset, UK) per 100 ml (85 V, 30 min). Gels were viewed and captured using ultra violet light, by a Bio-Rad image analyser with image lab software (Bio-Rad image lab, version 2.0.1), (Bio-Rad laboratories, Hertfordshire, UK).

2.4 PCR amplification for sequencing

PCR amplification was modified for the use of 454 Titanium FLX+ and MiSeq™ Sequencing due to the ability of 454 sequencing to provide longer fragment lengths in comparison to MiSeq sequencing. The V3-V5 region of the 16S rRNA gene was used for 454 sequencing while the shorter V2-V3 region was chosen for MiSeq analysis. These regions were chosen by the Wellcome Trust Sanger centre for their ability to distinguish between taxa using the available sequencing length. The move from 454 to MiSeq sequencing was carried out by the Wellcome Sanger centre due to the increased accuracy, superior number of returned reads and the reduction in costs, compared with 454 sequencing, in 2014, mid way through the project.

2.4.1 PCR amplification for Roche 454 Titanium FLX+™ sequencing

Barcoded primers targeting the V3 to V5 regions of the 16S rRNA gene were used to generate 16S rRNA gene amplicons for Lib-L454 Titanium sequencing from extracted DNA.

The primer 338F (5'-ACTCCTACGGGAGGCAGCAG) (MWG Eurofins, Ebersberg, Germany) and 926R (5'-CCGTCAATTCMTTTRAGT) (MWG Eurofins, Ebersberg, Germany) amplified

an approximately 550 base pair (bp) region of the 16S rRNA gene (5). 338F was modified at the 5' end with 454 adaptor B (5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG), while 926R was modified with 454 adaptor A (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG) followed by a unique 12 base barcode.

PCRs were performed in quadricate, with approximately 100 ng of template DNA used in each 20 µl reaction containing; 2 µl 10X AccuPrime™ PCR Buffer II (Invitrogen, Paisley, UK), which contained dNTPs, 0.08 µl AccuPrime™ Taq High Fidelity (Invitrogen, Paisley, UK), 0.2 µl of each 10 µM Primer to a final concentration of 0.1 µM. The thermocycling conditions were optimised to; initial denaturation 94°C for 3 min; 25 cycles of denaturation (94°C for 30 sec) annealing (55°C for 30 sec) extension (72°C for 2 min). PCR machine used throughout was the G-storm thermocycler (GRI, Braintree, Essex).The presence of DNA was verified by Tris-Borate-EDTA (TBE)(Severn Biotech, Worcester, UK) agarose gel electrophoresis on 1% w/v TBE gels containing 1 µl of 10 mg/ml ethidium bromide (Sigma-Aldrich, Dorset, UK) per 100 ml (85 V, 30 min). Gels were viewed and captured using ultra violet light, by a Bio-Rad image analyser with image lab software (Bio-Rad image lab, version 2.0.1), (Bio-Rad laboratories, Hertfordshire, UK).

2.4.2 PCR amplification for Illumina MiSeq™ sequencing

Barcoded primers targeting the V1 and V2 regions of the 16S rRNA gene were used to generate 16S rRNA gene amplicons for Illumina MiSeq, from extracted DNA.

16S rRNA gene was amplified using modified primers 27F (5'-AGMGTTYGATYMTGGCTCAG) (MWG Eurofins, Ebersberg, Germany) and 338R (5'-GCTGCCTCCCGTAGGAGT) (MWG Eurofins, Ebersberg, Germany) (6). The 5' end of the primer 27F was modified with a 29 base Illumina adapter, a 10 base forward primer pad and a 2 base linker sequence (5'-AATGATACGGCGACCACCGAGATCTACAC TATGGTAATT CC). The 5' end of 338R was modified with the reverse complement Illumina adapter, a 12 base Golay barcode, a 10 base reverse primer pad and a 2 base linker (5'-

CAAGCAGAAGACGGCATACGAGAT TCCCTTGTCTCC AGTCAGTCAG AA) (Underlined sequence represents an example Golay barcode). Primer pads are included to stop the formation of hairpins within the primers while primer linker regions are included to separate the 16S rRNA gene priming site from the unique barcode. Linker regions are required due to the variable barcode sequences which could potentially bind to the 16S rRNA gene sequence adjacent to the priming target, creating longer priming sites for some species. It is therefore important that the linker bases do not match the upstream 16S rRNA gene region for any known bacterial species. This ensures that priming occurs across the same number of bases for all species.

PCRs were performed in quadricate, with approximately 50 ng of template DNA used in each 25 µl reaction containing; 5 µl 5 X Q5 Buffer (New England Biolabs, Ipswich, UK), 1 µl Q5 High-fidelity polymerase (New England Biolabs, Ipswich, UK), 1.25 µl of each 10 µM Primer was added to a concentration of 0.5 µM, 0.5 µl 10 mM dNTPs. The thermocycling conditions were optimised by the Wellcomme Trust Sanger institute to; initial denaturation 98°C for 2 min; 25 cycles of denaturation (98°C for 30 sec) annealing (50°C for 30 sec) extension (72°C for 1 min 30 sec), final extension 72°C for 5 min.

The presence of product was confirmed by TBE agarose gel electrophoresis, as above (Chapter 2.4.1).

2.5 PCR clean-up

The four repeat PCRs carried out for sequencing were pooled and cleaned up using ethanol precipitation to remove salts from the extracted DNA. 0.3 volumes of 1M NaCl (Sigma-Aldrich, Dorset, UK) was added to the pooled PCR volume, two volumes (relative to the new combined sample including NaCl volume) of cold 100% ethanol were then added. The samples were then mixed by inversion and stored at -20^{o}C overnight. After incubation the samples were centrifuged at 4^{o}C at 15,000 X *g* for 20 min. The supernatant was removed

and 600 µl of cold 70% ethanol was added. The samples were then centrifuged at 4^{o}C at 15,000 X *g* for a further 15 min. The supernatant was removed and samples were left to air dry. Once dry samples were re-suspended in 30 µl of Tris-EDTA (TE) (Sigma-Aldrich, Dorset, UK) buffer and stored overnight at 4^{o}C.

2.6 Sequencing

All barcoded PCR products were quantified individually using a Qubit 2.0 fluorometer (Invitrogen, Paisley, UK) and combined into an equimolar mastermix prior to sequencing at the Wellcome Trust Sanger Institute (Cambridge, UK). Both 454 and MiSeq sequencing were performed at the Wellcome Trust Sanger Institute (Cambridge, UK).

2.7 Sequencing analysis

The Mothur sequencing analysis software platform was used to analyse the resulting data (7). Failed sequence reads, low quality sequence ends, tags and primers, were initially removed, followed by sequences shorter than 350bp (for 454 sequencing analysis) and 270bp (for Illumina sequencing analysis) and any sequences which included ambiguous base calls and homopolymeric stretches longer than 8 bases. Any non-bacterial ribosome sequences or chimeras were removed using Chimera Perseus software (8) or Chimera uchime (9) as implemented in Mothur. Sequences were assembled into operational taxonomic units (OTUs) (clustered at 97% identity), to give an approximation of species (10), and identified using the Ribosomal database project (RDP) training set (for version used see individual chapters) reference database (11). Representative sequences were used to give an appropriate species level identification for the OTU using NCBI megablast and any suspect OTUs, those that appeared in control samples as well as single sequences identified

as being unlikely to be found in humans (e.g. hyperthermophiles, methanotrophs and rhizosome associated bacteria) were removed.

2.8 Data and statistical analysis

Statistical analysis was performed in R, unless otherwise stated (12). Details of Chapter-specific statistical tests can be found in the corresponding methods sections. Details of statistical tests used throughout the study are detailed below.

2.8.1 Diversity measures

In order to assess the richness and evenness (equality of species distribution in an environment) of the bacterial community within each sample three complementary measurements of diversity were used. The indices chosen were species richness (*S**), Shannon-Wiener diversity index (*H'*) and Simpson's index of diversity (1-*D*), as described previously (13).

As explained by Magurran (2004) there are no perfectly unified diversity indices as different indices emphasize species richness and evenness to varying degrees (14). By using three metrics we are able to gain an indication of bacterial diversity while acknowledging that each measure has inherent bias. *S** indicates the total number of species but does not account for species abundance. *H'* reflects the diversity and evenness of species. *H'* is sensitive to changes in the frequency of common and less common species, however not rare species. 1-*D* is a measure of the probability that two species randomly selected from a sample will differ, however it is heavily weighted towards the most abundant species.

In order to account for sampling bias due to differing sampling depths (i.e. differing numbers of reads between sample, sample A may return 5,000 sequences while sample B returns 10,000) a randomised resampling method was used to examine differences in bacterial

diversity between samples, based on Solow (1993) (15). OTUs derived from a single sample were sub sampled, with replacement, to a uniform sampling depth, this allowed comparisons to be made between samples with different numbers of sequences. Diversity indices were calculated for each sample and a mean, median, standard error and 95% confidence intervals were calculated after 1000 iterations.

2.8.2 Similarity measures

Community compositions of samples were compared using two similarity measures Sørensen (number of shared species) and Bray-Curtis (accounts for the number and abundance of shared species). Both measures are considered to be highly effective (14). Sørensen is considered to be one of the most effective presence absence measures (16), while Bray-Curtis was found to be one of the few measures which could satisfy the following 6 criteria set out by Clarke and Warwick (2001): i) the value is one when two samples are the same; ii) the value is zero when samples have no common species; iii) a change of measurement does not affect the value of the index; iv) the value is unchanged by the inclusion of a species that do not occur in either sample; v) the inclusion of a third sample does not affect the similarity of the initial samples; vi) the index reflects differences in total abundance (17).

To overcome biases created by differences in sample size the randomised resampling method was adapted from that used for assessing diversity, in order to compare the similarity of equal subsamples from two original samples. Sørensen and Bray-Curtis similarity measures were calculated for each subsample in 1000 iterations and the mean, median, standard error and 95% confidence intervals were calculated.

2.8.3 Meta-analysis

Meta-analysis is a statistical method for comparing the results from different studies in order to identify patterns and relationships that may be missed in a single study. This technique is

widely used in clinical trials for the investigation of healthcare interventions (18), as well as being used by ecologists to compare different experimental studies (19). Overall patterns between studies are calculated by weighting means, therefore accounting for factors such as study size within the analysis. In this project, meta-analysis was used to assess the variation between samples by using the replication within the same sample as a single study. To calculate the variation with a sample the randomised resampling explained above (Chapter 2.8.1) was employed to give a mean and standard deviation. These values could then be used to provided a powerful tool to study between sample effects while accounting for the high variability between CF patients.

Meta-analysis was carried out in R, using the metafor package (20). The per-sample effect size was calculated using the standardized mean difference (SMD), widely used for the comparing two groups (treatment and control). This method uses the difference between sample means (m1i-m2i) divided by the pooled standard deviation (sd1i and sd2i). Expressed as:

$$SMD = \frac{m1i - m2i}{sd1i + sd2i}$$

Subsequently, the overall average effect size was calculated using a random-effects model using Hedges estimator (Hedges' *g*). Weighted estimation (with inverse-variance weights) was used to eliminate scale differences while incorporating the effect of between patient variations. Out puts were plotted in R, using ggplot2 (21).

2.8.4 Distribution abundance relationships (DAR)

It has been established that at the metacommunity level, a dichotomy exists within species, in order to investigate this species within the metacommunity it can be partitioned into those that are common, persistent and abundant, and rare, transient and less abundant. A direct, positive, relationship between species persistence and abundance has been observed in the study of taxa at a range of spatial scales, importantly this relationship has been observed

when within the CF airways (22, 23). As with animal species this distinction in the community composition has been found to be related to species permanence within the ecosystem, in this case the CF lung (24).

In this study we chose to use a partitioning method previously used in a study by Magurran and Henderson (2003) for the partitioning of fish species, from an estuarine environment (24). Persistence was described as the number of number of samples that contain the species of interest, while abundance was described as the total number of sequences from all samples. The log abundance was plotted against persistence to reveal a significant positive relationship. The most persistent and abundant organisms appearing in the upper quartile of the plot, i.e. in more that 75% of the total samples, these species were described as common, while all others were considered rare.

2.8.5 Analysis of variance (ANOVA)

Analysis of variance (ANOVA) is a statistical test of whether several means are equal, therefore providing a method of performing t-tests on more than two groups, without increasing the risk of type I error. After performing ANOVA tests, post-hoc Tukey honest significant difference (HSD) tests were then used to identify which groups in the sample were significantly different, $P<0.05$. This test allowed the comparison of OTU diversity with patients grouped by specific clinical factors.

2.8.6 Analysis of similarity (ANOSIM)

Analysis of similarity (ANOSIM) is a non-parametric test, used to test for significant differences between two or more groups (25). Results can be based on any distance measure, throughout this thesis comparisons were based on Sorensen and Bray-Curtis measure of similarity. This test allowed comparisons of community composition by clinical factors.

Similar to ANOVA, the ANOSIM is based on comparing distances within group to those between groups calculated;

$$R = \frac{r_b - r_w}{N(N-1)/4}$$

Where r_b= the mean rank of distances between groups and r_w= the mean rank of all distances. Significance is calculated using 10,000 permutations and pairwise ANOSIM tables are produced for *R* and *P*-values. Large positive *R* values signify dissimilarity between groups, while small values indicate groups are similar.

2.8.7 Similarity percentage (SIMPER)

Similarity percentage (SIMPER) analysis is used to assess the percentage contribution of taxa responsible for any observed differences between groups of samples (25). The overall significance of the difference in Bray-Curtis is assessed by ANOSIM.

2.8.8 Mantel and partial Mantel tests

Mantel tests are used to compare distance or similarity matrices. Using 9999 permutations the Pearson's correlation between two matrices are calculated resulting in an *r* value, between -1 and +1 and the corresponding significance. Partial Mantel tests are calculated in the same way however, the correlation is calculated while controlling for the effects of a third matrix;

$$r(AB.C) = \frac{r(AB) - r(AC)r(BC)}{\sqrt{1 - r(AC)^2}\sqrt{1 - r(BC)^2}}$$

where *r*(AB) is the correlation coefficient between A and B

Depending on the hypothesis, Mantel tests can be calculated as, lower tailed, two way or upper tailed. For the purposes of this project Mantel tests are calculated using a lower tailed

test. This hypothesised that more similar samples, in terms of community composition, would be closer in distance or similarity of clinical factors.

2.8.9 Mixed effects modeling

Mixed effects modelling is a statistical modelling technique that contains both fixed and random effects. These models are used in a wide variety of disciplines and are particularly useful for the analysis of longitudinal studies and studies which may have missing values. Mixed effects models can be represented as;

$$y = X\beta + Zu + \epsilon$$

y is a known vector of observations, with mean $E(y) = X\beta$; $\beta$is a unknown vector of fixed effects; u is an unknown vector of random effects, with mean $E(u) = 0$ and variance-covariance matrix $var(u) = G$; ϵ is the unknown vector of random errors and with mean $E(\epsilon) = 0$ and the variance $var(\epsilon) = R$.

Throughout this project mixed effects models were fitted in R, using the packages lmer (26), nlme (27), and GLMMADMB. Values of r^2 were calculated using the R package MuMIn. Details of all mixed effects models used are fully explained in the materials and methods sections of each chapter.

2.9 References

1. **Rogers, G. B., C. A. Hart, J. R. Mason, M. Hughes, M. J. Walshaw, and K. D. Bruce.** 2003. Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. J Clin Microbiol **41**:3548–3558.

2. **Nocker A, Sossa-Fernandez P, Burr MD, Camper AK.** 2007. Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol **73:**5111-5117.

3. **Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD.** 2008. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diag Microbiol Infect Dis **62:**133-141.

4. **Pitcher DG, Saunders NA, Owen RJ.** 1989. Rapid Extraction of Bacterial Genomic DNA with Guanidium Thiocyanate. Lett Appl Microbiol **8:**151-156.

5. **Muyzer G, de Waal EC, Uitterlinder AG**. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. App Environ Micorbiol **59**(3):695-700

6. **Lane DJ**.1991. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. Chichester, United Kingdom: John Wiley and Sons. 115–175

7. **Schloss PD, Gevers D, Westcott SL.** 2011. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLOS ONE **6**(12):e27310.

8. **Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ.** 2011. Removing noise from pyrosequenced amplicons. BMC bioinf **12:**38-55.

9. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:**2194-2200.

10. **Schloss PD, Handelsman J.** 2006. Toward a census of bacteria in soil. PLOS Comput Biol **2:**786-793.

11. **Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR.** 1996. The Ribosomal Database Project (RDP). Nucleic Acids Research **24:**82-85.

12. **The R Development Core Team.** 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

13. **Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, Lilley AK, Oliver A, Carroll MP, Bruce KD, van der Gast CJ.** 2013. Reducing bias in bacterial community analysis of lower respiratory infections. ISME J **7:**697-706.

14. **Magurran AE.** 2004. Measuring biological diversity. Blackwell Pub; Malden, MA, USA.

15. **Solow AR.** 1993. A Simple Test for Change in Community Structure. J Anim Ecol **62:**191-193.

16. **Southwood R, Henderson PA.** 2000. Ecological methods, 3rd ed. Blackwell Science, Oxford; Malden, MA, USA.

17. **Clarke KR and Warick RM.** 2001. Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. 2nd edition. PRIMER-E, Plymouth Marine Laboraotry, UK.

18. **Borenstein M.** 2009. Introduction to meta-analysis. John Wiley & Sons, Chichester, U.K.

19. **Marvier M, McCreedy C, Regetz J, Kareiva P.** 2007. A meta-analysis of effects of Bt cotton and maize on nontarget invertebrates. Science **316:**1475-1477.

20. **Viechtbauer W.** 2010. Conducting Meta-Analyses in R with the metafor Package. J Stat Softw **36:**1-48.

21. **Wickham H.** 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York.

22. **van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD.** 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. ISME J **5:**780-791.

23. **Rogers GB, Cuthbertson L, Hoffman LR, Wing PAC, Pope C, Hooftman DAP, Lilley AK, Oliver A, Carroll MP, Bruce KD, van der Gast CJ.** 2013. Reducing bias in bacterial community analysis of lower respiratory infections. ISME J **7:**697-706.

24. **Magurran AE, Henderson PA.** 2003. Explaining the excess of rare species in natural species abundance distributions. Nature **422:**714-716.

25. **Clarke KR.** 1993. Nonparametric Multivariate Analyses of Changes in Community Structure. Aust J Ecol **18:**117-143.

26. **Douglas Bates MM, Ben Bolker, Steve Walker.** 2014. Fitting Linear Mixed-Effects Models using lme4. R package version 1.1-7.

27. **Pinheiro J BD, DebRoy S, Sarkar D and The R Development Core Team.** 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117.

Chapter 3: Assessing the use of propidium monoazide

Assessing the use of propidium monoazide and its impact on the bacterial community within the CF lung

"Everything must be taken into account. If the fact will not fit the theory---let the theory go."

Agatha Christie, *The Mysterious Affair at Styles*

3.1 Introduction

In order to gain ecological insight into microbial communities it is important first to define the bacterial taxa present. By achieving this it becomes possible to assess community characteristics and predict the impact of perturbations on a system, through the use of a range of ecological measures, described in Chapter 2.8. This approach has been used widely to investigate microbial communities present in natural environments (1), with a move to adopting these techniques to investigate clinical systems in recent years. As technologies improve to allow this type of investigation, it becomes more important than ever to obtain the most accurate representation possible of the microbial community within a system.

Polymicrobial infections, caused by various combinations of bacteria, fungi, viruses and parasites, are commonplace in human disease (2). To effectively treat these infections it is important to increase our understanding of the roles bacteria play in disease prognosis. This is crucial when considering lower respiratory tract infections, particularly those associated with cystic fibrosis (CF). In CF, the lung damage sustained as a result of respiratory infections is a major driver of worsening lung function and respiratory failure in many individuals (3). As a consequence an improved understanding of the microbial community within the lungs may influence therapy and improve prognosis, by allowing targeted indiviualied therapies to be administered. However, in order for this to be achieved it is important to establish an accurate/unbiased picture of the CF lung community.

The use of culture independent methodologies for the identification and quantification of microbial species is increasing for a variety of reasons. These include their speed, reliability and ability to identify unusual and difficult to culture organisms as well as those in low abundance (4). Whilst these techniques avoid the selective biases associated with culture based microbiology they have their own limitations. For example; the ability to lyse cells, primer efficiency, contaminating DNA, as well as the intrinsic error rate involved in the use of sequencing technologies (5, 6). Another limitation that this study aims to address is the issue that standard PCR based analyses are unable to differentiate between live (or viable)

organisms, DNA associated with dead or damaged cells (including extracellular DNA). As such, this DNA may be amplified leading to overestimation of the abundance of living organisms. This can have implications for the analysis of microbial communities, as free DNA is known to persist for days or even weeks, depending on the environment, after cell death (7). Studies carried out on environmental samples found that DNA from dead or damaged cells is estimated to make up 70-74% of the total bacterial counts from marine sediment (8), while 76% of the total bacterial cells stained from a eutrophic river was not considered to be viable or active (9). Further, studies of clinical samples have revealed that if the issue of extracellular DNA or DNA from dead of damaged cells is not addressed, PCR based methodologies can result in false positives. For example, Panousis *et al* (2005) revealed a positive predictive value of 34%, when PCR was used for the diagnosis of infections in newly fitted prosthetic joints (10). This could lead to unnessecary treatments.

In the context of the CF lung, the exclusion of extracellular DNA and DNA from dead or damaged cells is immensely important as failure to identify effective treatments and therefore prolonged unnecessary treatment. The reduced ability to clear the airways (11), alongside the pressures from the host immune response (12) and antibiotic therapy (13) results in a highly perturbed system, where the bacterial community is under constant stress. In combination, these factors result in a build up of DNA not associated with viable cells within the lung (4, 14, 15).

The inclusion of extracellular DNA and DNA from dead and damaged cells in molecular analysis can have marked effects on community characterisation. This can lead to overestimation of bacterial density, a distortion of the actual relative abundance of species and the masking of rare, low abundant species by those in higher abundance. These importantly include the recognised pathogens, *Achromobacter xylosoxidans*, *Haemophilus influenzae*, and members of the *Burkholderia cepacia* complex (16). It may also mask any changes in community dynamics as a result of treatment or exacerbation, and it is therefore vital to exclude non-viable DNA from analysis. Consequently, this study uses propidium

monoazide (PMA) treatment to investigate the impact of extracellular DNA and DNA from dead or damaged cells on the bacterial community composition of CF sputum samples.

PMA is a membrane impermeable dye that covalently binds to DNA in the presence of light effectively reducing the amount of extra cellular DNA and DNA from dead or damaged cells in a sample (7). PMA treatment has been shown to be effective, in a range of contexts, for assessing viable microbes using molecular techniques. For example, Nocker *et al* (2007) used end-point PCR and denaturing gradient gel electrophoresis (DGGE) to assess the efficacy of PMA treatment on the bacterial species detected in environmental samples. PMA treated samples which had been previously spiked with known quantities of heat killed *Escherichia coli* O157:H7 showed PMA could successfully suppress DNA signals (17). While Bae *et al* (2009) found that PMA treatment could be effectively used to quantify viable *Bacteroidales* cells from human faeces and wastewater influent and effluent (18). Further Kralik *et al* (2010) used similar strategies to ascertain that PMA treatment, at a concentration of 25 µM for 5 min, followed by 2 min of light exposure, prior to qPCR could help accurately quantify viable *Mycobacterium avium* subsp. *paratuberculosis* as an alternative to culture (19).

Importantly for this investigation, the use of PMA has been assessed in the context of CF airway infections. In 2008, Rogers *et al*, used T-RFLP in investigate the effect of PMA treatment on the bacterial community within CF sputum. This study highlighted the significant contribution of non-viable bacteria in the CF lung to the signal obtained by T-RFLP analysis and the importance of avoiding this bias in community analysis (4). Further work, was also published by Rogers *et al*, in 2010, and revealed that failure to remove the influence of non-viable organisms could result in the inability to identify statistically significant changes in bacterial load, particularly that of *Pseudomonas aeruginosa* (15). In 2013, PMA was used to assess community changes using qPCR and 454 high-throughput pyrosequencing, revealing that although there were no significant changes in the overall diversity and community membership, when samples were partitioned into common and rare, PMA treatment highlighted rare community diversity which was found to be obscured by the most common

and abundant species (16). Importantly for clinicians and patients, this work showed that the detection of recognised pathogens, particularly *Achromobacter xylosoxidans* (undetected to satellite), *Haemophilus influenzae* (undetected to satellite), members of the *Burkholderia cepacia* complex (undetected to core) and *Mycobacterium* sp. (undetected to satellite), was significantly influenced by PMA treatment (16). Failure to identify these species may have considerable influence on treatment decisions and therefore patient outcomes.

In this study 16S rRNA gene pyrosequencing was used to assess how the bacterial community is affected by pre-treatment of samples with PMA. This research was undertaken to critically evaluate the use of PMA for the exclusion of free DNA and DNA from dead or damaged cells from the analysis of the bacterial community from CF lung infections, while allowing modelling techniques for the analysis of bacterial community changes to be evaluated. Finally, this study allowed the results from previous studies into the effect of PMA treatment to be investigated and confirmed. By assessing the use of PMA and its impact on the bacterial community, this study can be used to underpin the results obtained throughout this research project, allowing confidence that the results obtained reflect the true bacterial community.

3.2 Methods

3.2.1 Sample collection

Sputum samples were collected from 42 adult CF patients attending the regional Cystic Fibrosis Centre in Southampton General Hospital, under full ethical approval from Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26). Samples were collected and stored at -80°C until processing.

3.2.2 DNA extraction and Pyrosequencing

Sterile scalpels were used to divide frozen samples into two equal portions, prior to a sputum wash being performed as described in Chapter 2.3.1. The first portion was treated with PMA to exclude free DNA and DNA from non-viable cells from analysis via the method described in Chapter 2.3.2. The second half of the sample was left untreated prior to DNA extraction. DNA extraction was then performed on all samples as previously stated in Chapter 2.3.3. Following the method outlined in Chapter 2.4.1, bacterial Golay barcode encoded 454 FLX Titanium amplicon pyrosequencing was performed using the primer 338F (5'-ACTCCTACGGGAGGCAGCAG) and 926R (5'- CCGTCAATTCMTTTRAGT). 454 pyrosequencing using the Lib-L kit was performed at the Wellcome Trust Sanger Institute, Hinxton, UK. Resulting data was analysed using the Mothur sequencing analysis platform as previously described in Chapter 2.7 (20). The sequence data reported in this paper have been deposited in the European Nucleotide Archive under Study Accession Number; ERP007059, and Sample Accession Number; ERS551400. The relevant barcode information for each sample is shown in Table A3.1.

3.2.3 Statistical analysis

To determine whether sample size was large enough to effectively assess the bacterial diversity within each sample, Mothur was used to create intra-sample rarefaction curves without replacement (20). All further statistical analysis was performed in R version 3.1.1(2012-07-10) (21).

Changes in bacterial diversity were assessed using three complementary measures; species richness (*S**), Shannon-Wiener (*H'*), and Simpson's (1-*D*) indices of diversity as described previously, Chapter 2.8.1. To avoid potential bias all measures were calculated using a uniform, randomised resampling, see Chapter 2.8.1 (16). Diversity measures were re-sampled to the lowest number of sequence reads (*n*=202).

Meta-analyses were used to summarise the effect of PMA treatment on measures of bacterial diversity and composition using Hedges estimator of effect size, as explained in Chapter 2.8.3. The R package Metafor, was used to compare treated to untreated samples

by summarising the overall effect of the 36 paired samples, each patient was treated as a separate independent study (22). The overall effect of treatment was calculated using the standardised mean difference for each sample was calculated using hedges *g*.

Species were partitioned using a variety of criteria, explained below, to allow investigation of the effect of PMA treatment on different aspects of the bacterial community. Firstly, species were partitioned into common, the most persistent and abundant, and rare, transient and less abundant, using a distribution abundance relationship, described by Magurran and Henderson (2003) (23). Common species were described as those present in more that 75% of the total samples, while all others were considered to be rare. Bacterial species were then partitioned according to oxygen tolerance; the aerobic species group consisted of all aerobic species, facultative anaerobes and microaerophiles. The anaerobic species group contained only strict anaerobes. Finally members of the oral microbiota were identified using the human oral microbiome database (HOMD, http://www.homd.org).

The similarity of treated samples were compared to untreated in order to assess the changes community composition as a result of PMA treatment using Bray-Curtis measure of similarity (accounting for both presence/absence and abundance of species), as described in Chapter 2.8.2. To avoid potential bias Bray-Curtis measure of similarity was calculated using a modified version of the uniform resampling method previously used (*n*=202), see Chapter 2.8.2 for full details (16).

Mixed effects models, using the R package lme4 were used throughout the analysis of this data to allow comparisons between treated and untreated samples using a variety of diversity and composition measures as well as using different partitioning criteria to gain an overall understanding of the effect of PMA treatment in the bacterial community (24). All models used patient as a random effect in order to account for between patient variability. r^2 values were calculated using the MuMIn package in R (21).

3.3 Results

A total of 260087 bacterial sequences (mean = 2653.9 ± SD 2196.7, *n*=96), were generated with 152818 sequences from non-PMA samples (mean = 3118.7 ± SD 2322, *n*=49) and 107269 sequences from PMA samples (mean = 2189.1 ± SD 1979.6, *n*=49).

The diversity and composition of bacterial communities within each of the paired samples (non-PMA *n*=42, PMA *n*=42) was assessed by 16S rRNA gene 454 pyrosequencing. Rarefaction curves were used to test whether the number of sequences per sample was sufficiently large to capture a robust cross-section of the bacterial diversity (Figure 3.1.1, Figure 3.1.2). To this end, 6 paired samples, from patient 5, 26, 28, 32, 40 and 55, were removed due to insufficient sampling depth, indicated by rarefaction curves not reaching an asymptote. This resulted in, no PMA treatment *n*=36 and PMA treatment *n*=36, (Total samples *n*=72), samples being included in the study for further analysis. Following removal of these samples, a total of 215649 bacterial sequence reads (mean = 2995.11 ± SD 2012.41, *n*=72) identifying 41 genera and 98 bacterial species remained, a complete list is shown in Table A3.2. No difference was observed between the PMA and non-PMA samples after the removal of samples that did not have the required sampling depth.

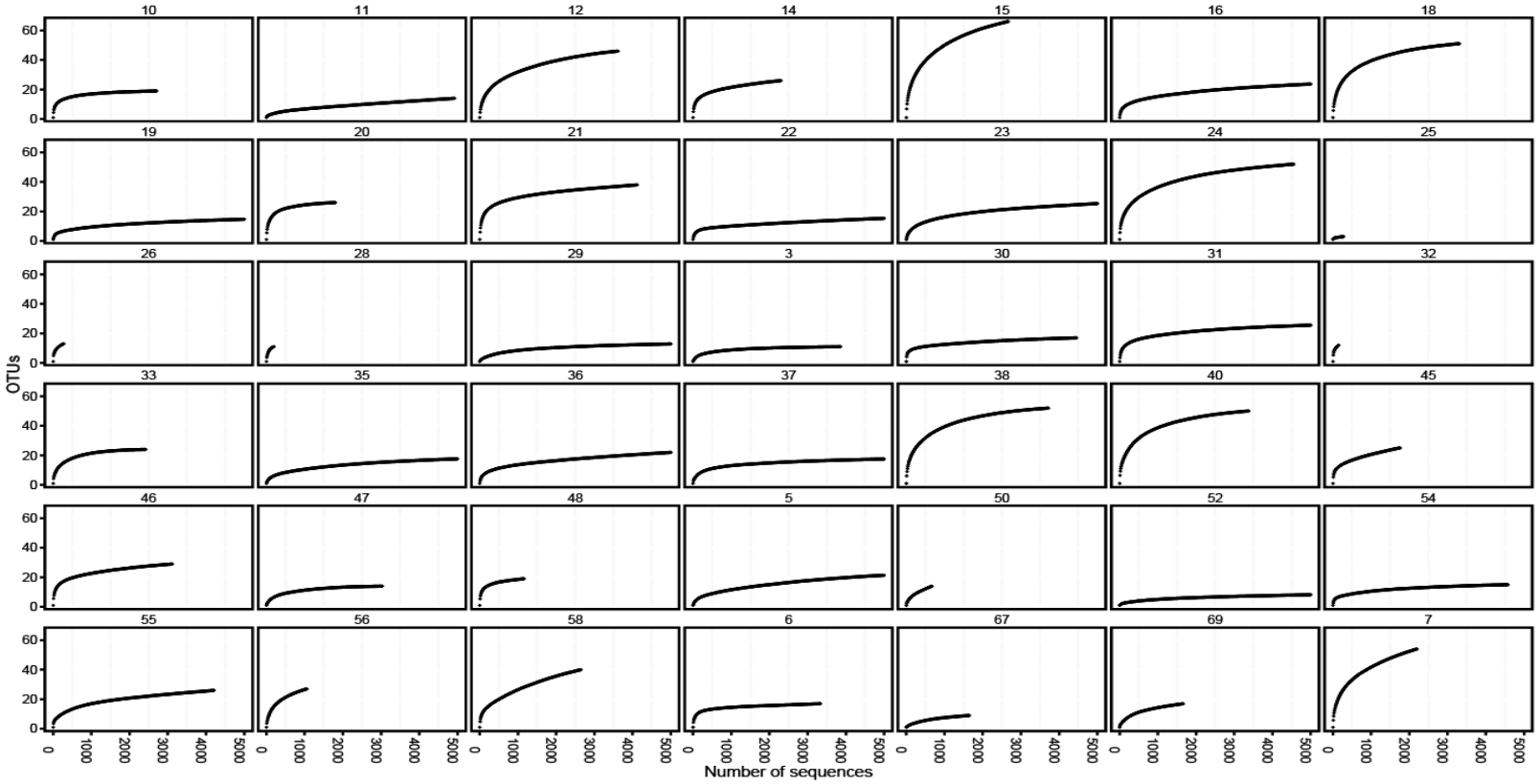


Figure 3.1.1 Rarefaction curves for non-PMA treated sputum samples.

Intra-sample rarefaction curves were generated in Mothur, using a resampling without replacement approach. Samples which did not reach or approach asymptote were removed from further analysis along with the paired PMA treated sample.

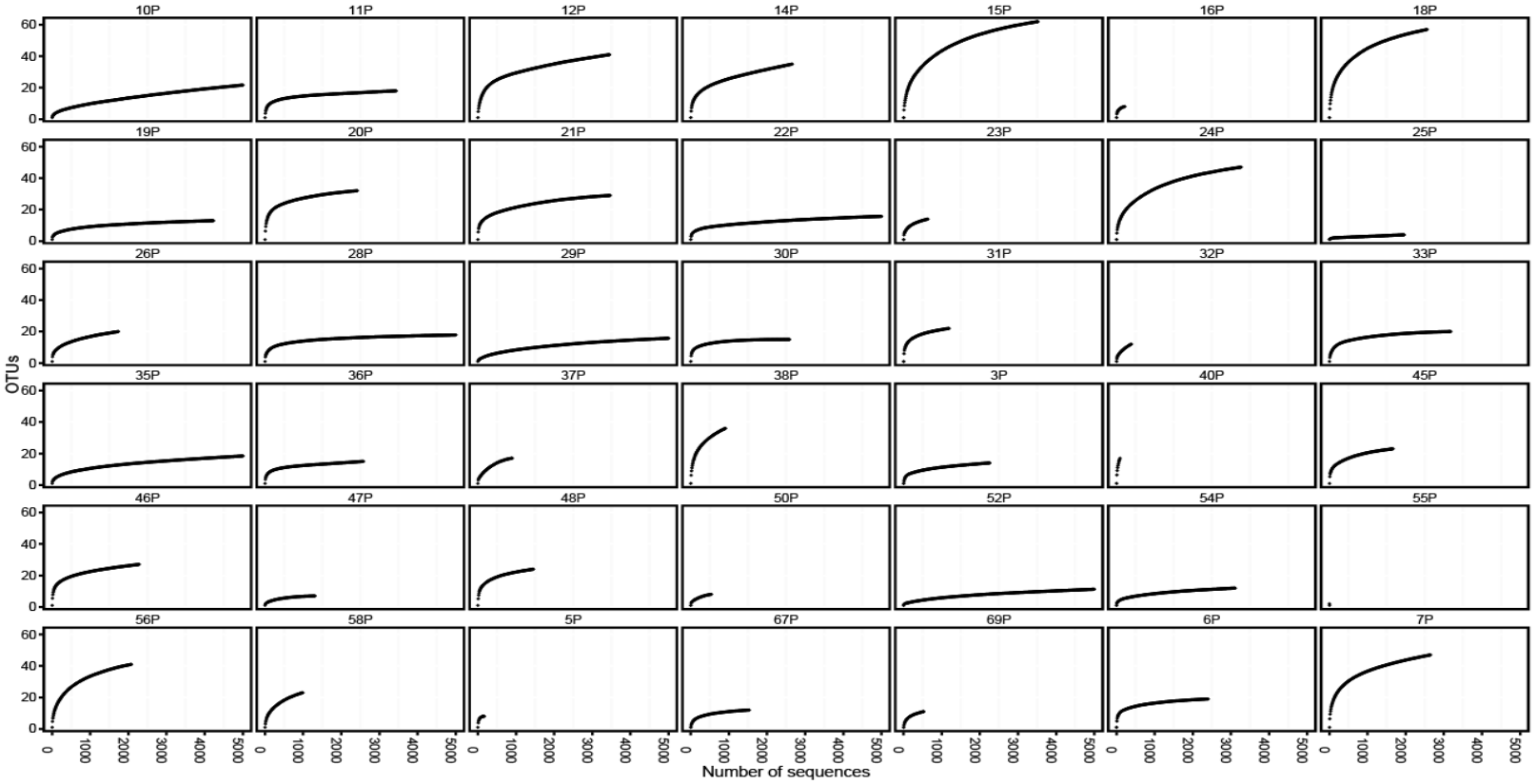


Figure 3.1.2 Rarefaction curves for PMA treated sputum samples.

Intra-sample rarefaction curves were generated in Mothur, using a resampling without replacement approach. Samples which did not reach or approach asymptote were removed from further analysis along with the paired non-PMA treated sample.

3.3.1 Diversity

As a result of varying levels of sampling depths obtained, a feature of high throughput sequencing, a randomised resampling method (with a uniform sub-sample size of n=202 sequences, equating to the smallest number of sequences recovered), as described in Chapter 2.8.1, was applied to the samples to reduce potential bias in the comparison of community diversity. Three recognised metrics of community diversity were applied to the data, species richness (*S**), Shannon-Wiener index (*H'*), and Inverse Simpson's index (1-*D*). All diversity measures were found to be highly variable between individual samples, regardless of treatment. As shown in Figure 3.2, meta-analysis revealed both significantly positive and significantly negative effects of PMA treatment for each individual sample. However, the overall effect of each diversity measure, calculated using Hedges *g*, revealed no overall significant changes, indicated by the confidence intervals crossing 0, in any of the 3 diversity measures, shown in detail in Figure 3.3. Mixed effect models confirmed the findings of the meta-analysis, *S**, P=0.542, r^2=0.0004, *H'*, P=0.070, r^2=0.008, 1-*D*, P=0.051, r^2=0.012. Changes in overall species richness were observed between the treated and non-treated samples, with 81 species observed in the non-treated samples compared with 91 in the treated however, this was not significant when between patient variability was accounted for.

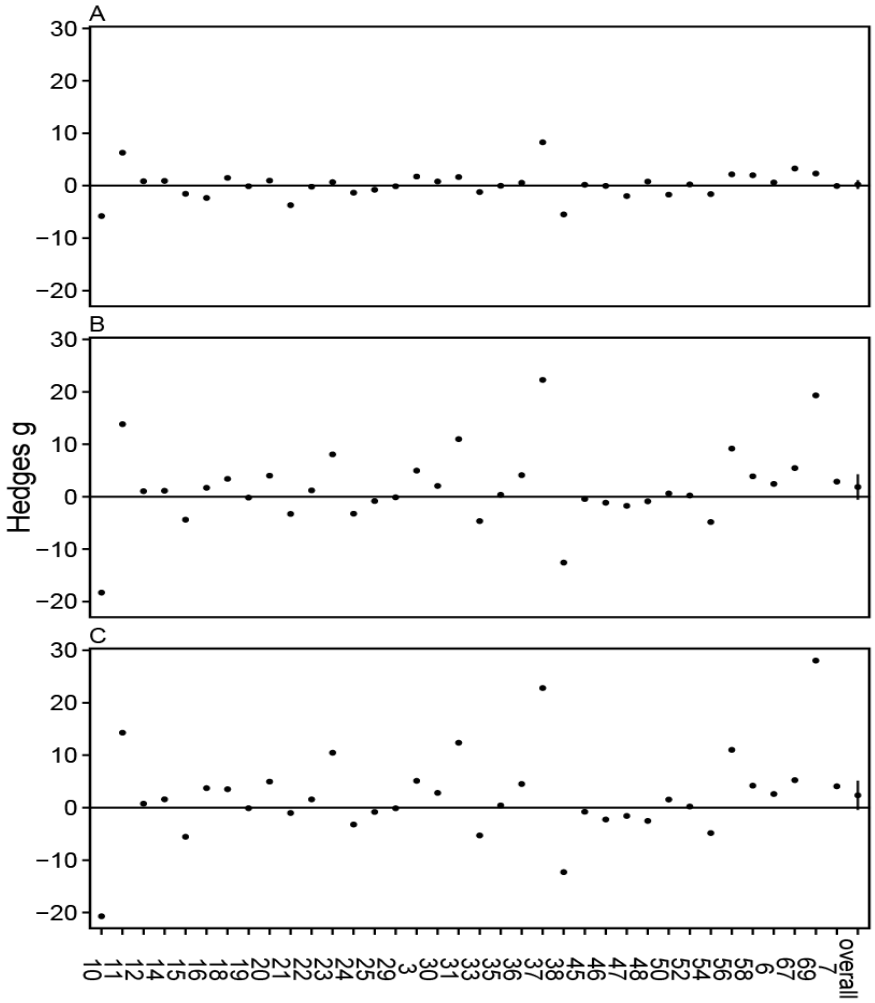


Figure 3.2 The effect of PMA treatment on the bacterial community diversity using weighted mean differences.

A) species richness (S*), B) Shannon-Wiener index (*H'*) and C) Inverse Simpson's (1-*D*). The overall effect of PMA treatment on each diversity measure was calculated using Hedges' *g* effect size measure, error bars represent 95% confidence intervals of the effect size (*n*=36). Negative values indicate lower diversity in PMA treated samples. Error bars crossing zero indicate no significant effect of PMA treatment.

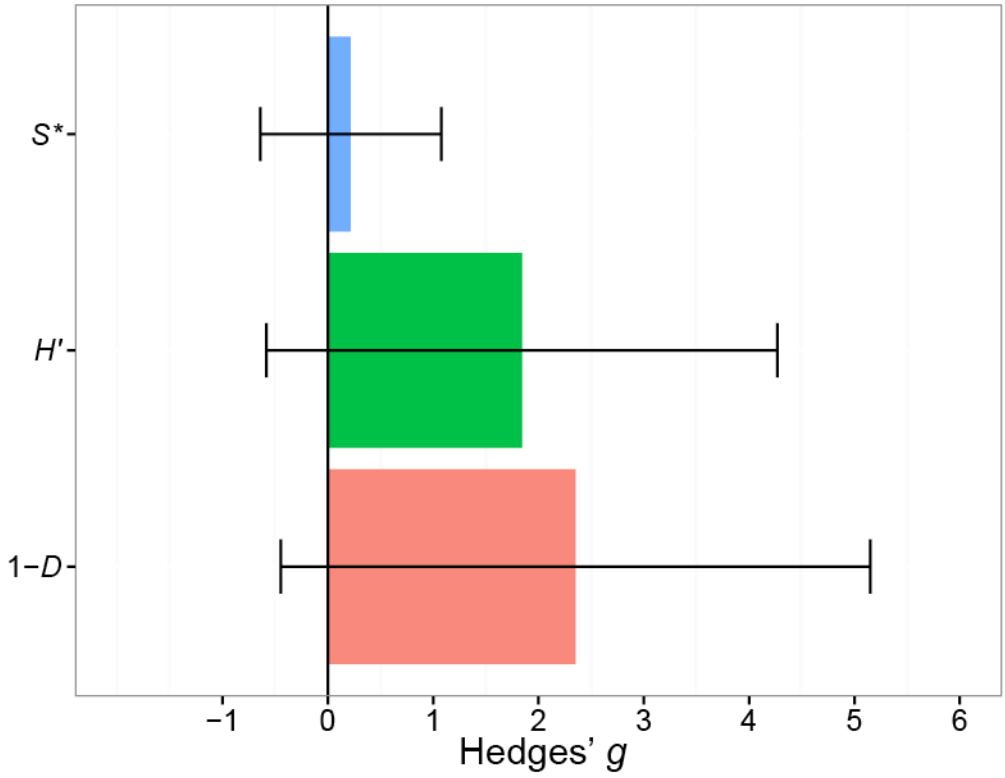


Figure 3.3 The overall effect of PMA treatment on the diversity measure.

Species richness (S*), Shannon-Wiener index (*H'*) and Inverse Simpson's (1-*D*) using Hedges' *g* effect size measure, error bars represent 95% confidence intervals of the effect size (*n*=36). Error bars crossing zero indicate no significant effect of PMA treatment.

A distribution abundance relationship (DAR), was employed to partition the species in those considered to be common (both persistent and abundant) and those considered to be rare (transient with low abundance). The log species abundance was plotted against the sample persistence (the number of samples in which a particular species appears) showing a significant positive relationship for both the untreated (n=36, $F_{(1,83)}$=310.6, P<0.001), and treated (n=36, $F_{(1,89)}$=243.6, P<0.001) samples. This direct relationship between species persistence and abundance indicates a coherent metacommunity. Species present in more than 75% of the samples were denoted as common, all others were deemed rare. As shown in Figure 3.4, the DAR revealed 5 species present in more than 75% of both the treated and untreated samples*, Prevotella melaninogenica, Veillonella dispar, Streptococcus pneumoniae, S. parasanguinis* and *Pseudomonas aeruginosa*. *Gemella morbillorum* was categorised as common in the untreated samples and rare in those exposed to PMA treatment.

Mixed effect models were used to investigate the effect of PMA treatment on the species identified as common. These changes were found to be variable by species, a significant positive effect of PMA treatment was observed in the relative abundance of the OTUs, *S. pneumoniae* (P<0.001, r^2=0.07) and *S. parasanguinis* (P<0.001, r^2=0.05). While PMA treatment was found to have a significant negative effect on the abundance of; *P. aeruginosa* (P=*0.004*, r^2=0.03). No significant change in relative abundance was observed for *P. melaninogenica* (P=0.024, r^2=0.02) *or V. dispar* (P=*0.03*, r^2=0.04). *G. morbillorum* was the only species to be classified as both common and rare, being common in the untreated samples but rare in the treated, despite this no significant difference in abundance was observed (P=0.058, r^2=0.01). Of the rare species present in more than 50% of the samples (but <75%), PMA was found to result in significant increase in the abundance of; *Bacteroides acidifaciens* (P=0.01, r^2=0.03), *Shuttleworthia satelles* (P=0.04, r^2=0.01), *Granulicatella adiacens* (P<0.001, r^2=0.05), and *Porphyromonas catoniae* (P=0.006, r^2=0.06).

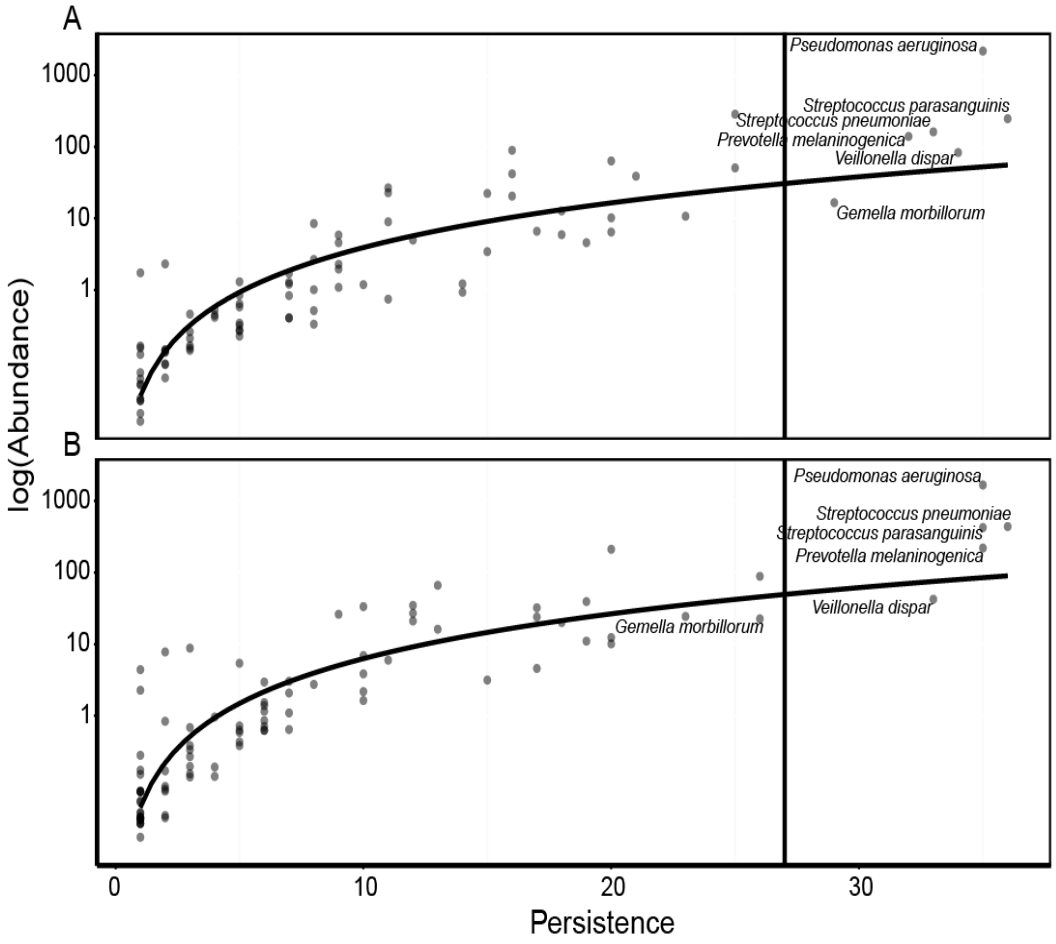


Figure 3.4 Persistence abundance curves.

A) Untreated samples (n=36, $F_{(1,83)}$=310.6, P<0.001), B) PMA treated samples (n=36, $F_{(1,89)}$=243.6, P<0.001). Common species were defined as those that fell in the upper quartile (right of the vertical line), all species that fell below the line were considered to be rare, These OTUs are shown in Table S2.

3.3.2 Aerobic and anaerobic species

Recent studies have revealed a steep oxygen gradient within the mucus of the CF lung and as a result anaerobic species have been repeatedly isolated from the lung of CF patients through both culture and culture independent methodologies (25-27). To investigate the contribution of anaerobes to the bacterial community within the CF lung all species were partitioned into either aerobic or anaerobic species. All species able to grow in aerobic conditions, including facultative anaerobes and microaerophiles, were considered aerobic. Only strict anaerobes were partitioned into the anaerobic group. Species identified as anaerobes are indicated in Table A3.2.

No significant change in the relative abundance of aerobes (n=44, P=0.813, r^2<0.001) and anaerobes (n= 54, P=0.655, r^2<0.001) was found when samples were treated with PMA. However, when diversity measures were applied there was a significant increase in Simpson's diversity index for the aerobic group (P=0.036, r^2=0.021) when samples were treated with PMA. This was not observed for S^* (P=0.121, r^2=0.007) or H' (P=0.087, r^2=0.013). Furthermore this change was not observed in the anaerobic species; S^*, P=0.108, r^2=0.005, H', P=0.763, r^2<0.001, 1-D, P=0.606, r^2=0.002. This shows that more rare aerobic OTUs are identified with the use of PMA, suggesting that when free DNA is removed from the sample a more complete picture of rare OTUs can be observed.

3.3.3 The oral microbiota

Using the strict criteria above, both aerobic and anaerobic species are found to make up the oral microbiota. Several studies have investigated the presence of members of the human oral microbiota in the CF lung, establishing that their presence is unlikely to be due to contamination, but true members of the bacterial community (25, 28). Therefore, oral species present within the CF lung represent potential pathogens that may be clinically important. In this study 75 of the bacterial species identified were known to be commonly present in the oral cavity, Table 3.1 (29). All species partitioned as common were identified as members of

the oral microbiota. Treating samples with PMA resulted in a significant decrease in the abundance of the total oral species detected (P<0.001, r^2=0.02) and S* (P=0.026, r^2=0.09), however, no changes in the community diversity were observed; S*, P=0.185, r^2=0.003, H', P=0.755, r^2<0.001, 1-D, P=0.614, r^2<0.001.

3.3.4 Community composition

In order to compare changes in community composition between treated and untreated samples, the Bray-Curtis measure of similarity (S_{BC}, which accounts for the number and abundance of species present in each community and those that are shared) was used. To minimise possible bias due to sampling depth, S_{BC} was calculated using randomised resampling to assess the similarity of treated to untreated samples. High variability was observed between samples with a range of similarities between 0.99 and 0.11, as shown in Figure 3.5.

To investigate possible explanations for the variation in similarity observed, the results of S_{BC} were compared to values of dominance. The Berger-Parker measure of dominance, the proportional abundance of the most abundant species present in the community (30), was calculated showing a significant decrease in dominance when samples were treated with PMA (P=0.028, r^2=0.02). To investigate any relationship between dominance and changes in similarity, Berger-Parker was plotted against S_{BC}. A significant positive relationship was found between change in similarity and dominance (P<0.001, r^2=0.34). Samples more highly dominated by a single species were found to show less change in similarity with PMA treatment than those with a more diverse community, Figure 3.6. Dominance was also found to have a significant relationship with species richness (P<0.001, r^2=0.48), where less diverse communities were found to be more highly dominated by one or a small number of species.

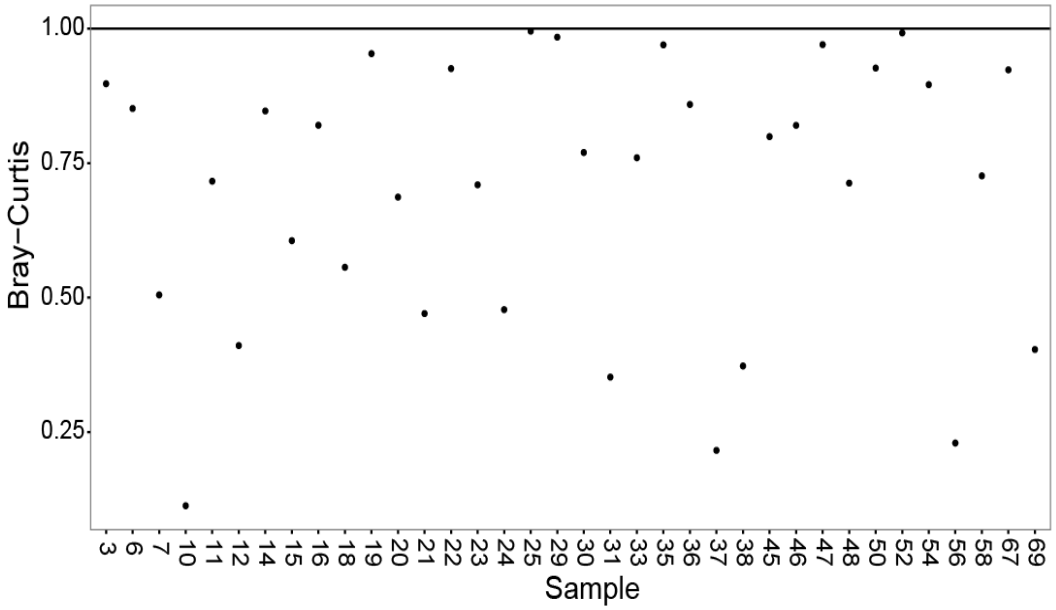


Figure 3.5 The effect of PMA treatment on community composition using the Bray-Curtis measure of similarity.

Points represent the Bray-Curtis similarity between non-treated and PMA treated samples for each individual patient. The closer the given value is to 1 the more similar the community is to the original untreated sample community. This suggests some patients show very few changes in community composition with PMA treatment while other show considerable changes in both the OTUs identified and their abundance, leading to values closer to 0.

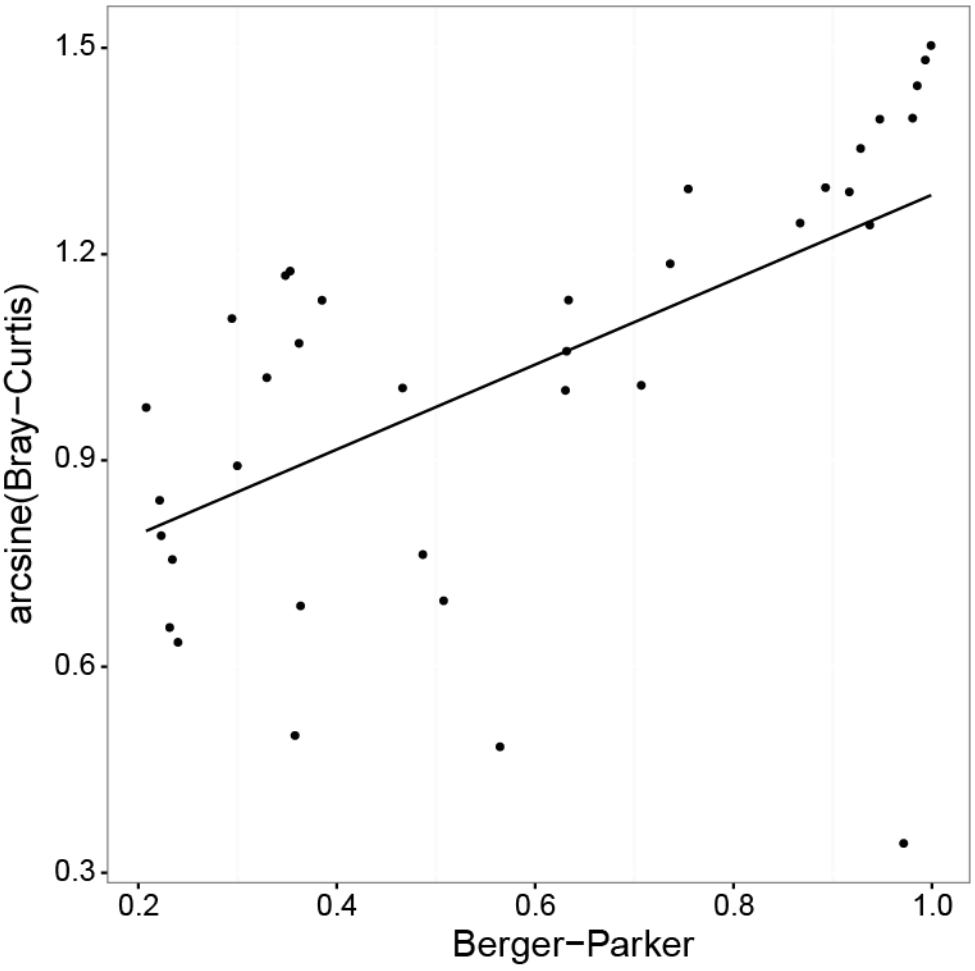


Figure 3.6 The relationship between community similarity (Bray-Curtis) and dominance (Berger-Parker).

The difference in similarity, between non-treated and PMA treated samples using Bray-Curtis similarity, plotted against the dominance measured by Berger-Parker dominance measure. A linear model has been fitted to the data (P<0.001, r^2=0.34) in order to highlight the significant relationship between similarity and dominance. An arc-sine transformation was applied to the Bray-Curtis measure in order to make the data normal in order to correctly fit the model without violating assumptions.

3.4 Discussion

As highlighted in the introduction (Chapter 1.5), the lower respiratory tract can be a challenging environment for bacterial species to colonise and persist in. Despite the lack of an effective mucociliary escalator in patients with CF, bacteria colonising the CF lung, continue to be subjected to host immune responses as well as a wide variety of antimicrobial therapies (16). Whilst some cells may be capable of growth in this context, the host immune response and use of antimicrobials mean that only a certain percentage of cells entering the lower respiratory tract will be viable. By conventional molecular analysis, DNA from dead cells, lysed cells and otherwise non-viable cells still have the potential to contribute signal to culture-independent studies. As such, the persistence of extracellular DNA and that from dead or damaged microbial cells under this perturbed system makes it challenging to define the bacterial community using culture independent technologies. This study aimed to establish the effect of pre-treating sputum samples, from patients suffering from CF, with PMA.

Several previous studies have observed high variability in the bacterial communities between CF patients (14, 31, 32). In order to account for this high degree of inter-individual variation several methods are available. In this study, meta-analysis, which allows an overall effect size to be calculated from weighted mean differences, is compared to mixed effect modelling, which can be used to model any changes between the groups while accounting for between patient differences. Both methods showed no significant overall change in diversity with the use of PMA. However, the use of the meta-analysis approach supported the findings of Rogers *et al* (2013) suggesting that without the use of PMA the diversity of the local bacterial community in individual samples can be under or over estimated.

Previous studies have established that there is a divide in the metacommunity, between the common and locally abundant species and those that are rare or transient, within the CF lung (32, 33). As a result failure to make the distinction between groups can result in important aspects of the species abundance distributions being neglected. In such studies,

the core-satellite group approach was established using a Poisson distribution to partition species into groups. This method allowed samples to be partitioned into those that were randomly (satellite) and non-randomly (core) distributed. However, due to a quirk in this analysis some species with low persistence across the dataset could be assigned as a core species due to their high abundance in a small number of samples. To overcome this, a distribution abundance relationship was employed to partition and allow only the species appearing in more than 75% of the samples to be considered common. This distribution showed a direct relationship between persistence and abundance, indicating a coherent metacommunity. By partitioning identified species into common and rare groups it was observed that the use of PMA treatment had a significant effect on the abundance of *S. pneumoniae*, *S. parasanguinis* and importantly the recognised pathogen *P. aeruginosa*. Significant changes in the abundance of these species could have marked consequences for downstream analysis of sequencing, as important community changes as a result of treatment or other perturbations may be masked. This could also have implications for treatment decisions that may be informed by misrepresentative data.

Rogers *et al* (2013) observed changes in species distribution between the treated and non-treated samples, importantly the movement of recognised pathogens were observed between the core and satellite groups (16). However, in this study only *Gemella morbillorum* moved between the common and rare groups with PMA treatment. No movement of recognised CF pathogens was observed, which may be due to the more stringent partitioning applied to the community, or alternatively, a result of the highly *P. aeruginosa* dominated communities observed in this study.

It has been well established that anaerobic species make up an important part of the bacterial community within the CF lung (25-27). As such, all identified species were partitioned into anaerobes (strictly anaerobic species) and aerobes (aerobic, facultative anaerobes or microaerophiles) in order to establish the effect of PMA on the relative abundance and community diversity of each species group. While it was observed that treatment with PMA had no overall effect on the anaerobic community diversity, a significant

increase (P<0.05) in Simpson's index of diversity was shown in the aerobic community. Furthermore, the fact that there was no change in the other diversity measures used suggests that the community shift observed was not simply to do with a change in richness or abundance, but rather a combination of the two factors leading to an increase in community diversity. One explanation may be that the antibiotic treatments used targeted the most abundant aerobic species, thereby allowing more of the rare community to be identified.

Questions about whether the detection of members of the oral microbiota is due to sampling contamination rather than their presence within the lung community have been raised in previous studies (26, 28, 32, 34). Despite the use of PBS washes to remove saliva, and therefore possible oral contamination, the majority (n=75) of the species identified within our sample set are known to be present in the human oral cavity. This strongly suggests these species are true (and important) members of the lower respiratory bacterial community, perhaps colonising as a result of microaspiration of microbes from the oral cavity into the lower respiratory tract. While no change in diversity was observed, there was a significant decrease in the relative abundance of the members of the oral community, implying that PMA treatment changes the community composition allowing less abundant aerobic members of the community to be identified.

Bray-Curtis measure of similarity was used to investigate changes in community composition. PMA treated samples were compared to the untreated sample in all cases, revealing a high level of variation between samples. By comparing these results to those generated using the Berger-Parker sample dominance measure, samples revealed to be more highly dominated by a single species showed less change in community composition with PMA treatment. Patients dominated by a single species were found to have less diverse bacterial communities in terms of species richness. This would explain the conserved nature of dominated communities, where the dominant species out competes less abundant taxa for space and resources. As such, similar patterns have been observed when the diversity and community composition of grasslands was investigated (35). Therefore, the effect of PMA

treatment on the diversity and community composition of highly dominated samples is expected to be significantly less than those with more diverse communities.

The CF lung contains a wide range of species, making up highly variable communities under constant pressure from a wide range of immune and therapeutic factors (36, 37). These perturbations within the lung create a dynamic community that is constantly evolving. To investigate how the community is affected by perturbations, it is important to be able to differentiate between viable organisms and DNA from dead or damaged cells. This is particularly important when considering the effect of antibiotic treatment regimes, which may be masked by the presence of DNA from dead or damaged cells, resulting in skewed results and ultimately failure to assess the effectiveness of treatment.

In this study it the inclusion of DNA from dead or damaged cells prevents accurate characterisation of the bacterial community by leading to an over or under estimation of diversity and significant changes in the relative abundance of the most common species, as well as those characterised as members of the oral microbiota. Analysis of the viable community can allow the identification of rare bacterial species that would otherwise be masked by DNA from non-viable sources. This study also highlighted the huge variation observed in bacterial community composition between patient samples. By using mixed effect models, which allow this variation between patients to be investigated, the effect of PMA treatment could efficiently be calculated for a range of ecological measures.

From a clinical perspective, the use of PMA treatment for the exclusion of DNA from dead and damaged cells appears increasingly vital to provide accurate information on treatment progression and effectiveness. Therefore PMA treatment will be incorporated into the methodologies used to prepare samples for culture independent sequencing techniques throughout this dissertation.

3.5 References

1. **Ager D, Evans S, Li H, Lilley AK, van der Gast CJ.** 2010. Anthropogenic disturbance affects the structure of bacterial communities. Environ Microbiol **12:**670-678.

2. **Brogden KA, Guthmiller JM, Taylor CE.** 2005. Human polymicrobial infections. Lancet **365:**253-255.

3. **Davis PB, Drumm M, Konstan MW.** 1996. Cystic fibrosis. Am J Resp Crit Care Med **154:**1229-1256.

4. **Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD.** 2008. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diagn Microbiol Infect Dis **62:**133-141.

5. **von Wintzingerode F, Gobel UB, Stackebrandt E.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev **21:**213-229.

6. **Kunin V, Engelbrektson A, Ochman H, Hugenholtz P.** 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol **12:**118-123.

7. **Nocker A, Cheung CY, Camper AK.** 2006. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. J Microbiol Methods **67:**310-320.

8. **Luna GM, Manini E, Danovaro R.** 2002. Large fraction of dead and inactive bacteria in coastal marine sediments: comparison of protocols for determination and ecological significance. Appl Environ Microbiol **68:**3509-3513.

9. **Freese HM, Karsten U, Schumann R.** 2006. Bacterial abundance, activity, and viability in the eutrophic River Warnow, northeast Germany. Microb Ecol **51:**117-127.

10. **Panousis K, Grigoris P, Butcher I, Rana B, Reilly JH, Hamblen DL.** 2005. Poor predictive value of broad-range PCR for the detection of arthroplasty infection in 92 cases. Acta Orthop **76:**341-346.

11. **Döring G, Gulbins E.** 2009. Cystic fibrosis and innate immunity: how chloride channel mutations provoke lung disease. Cell Microbiol **11:**208-216.

12. **Elizur A, Cannon CL, Ferkol TW.** 2008. Airway inflammation in cystic fibrosis. Chest **133:**489-495.

13. **Ratjen F.** 2001. Changes in strategies for optimal antibacterial therapy in cystic fibrosis. Int J Antimicrob Agents **17:**93-96.

14. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Kehagia V, Jones GR, Bruce KD.** 2005. Bacterial activity in cystic fibrosis lung infections. Respir Res **6:**49.

15. **Rogers GB, Marsh P, Stressmann AF, Allen CE, Daniels TV, Carroll MP, Bruce KD.** 2010. The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. Clin microbiol and infect **16:**1656-1658.

16. **Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, Lilley AK, Oliver A, Carroll MP, Bruce KD, van der Gast CJ.** 2013. Reducing bias in bacterial community analysis of lower respiratory infections. ISME J **7:**697-706.

17. **Nocker A, Sossa-Fernandez P, Burr MD, Camper AK.** 2007. Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol **73:**5111-5117.

18. **Bae SW, Wuertz S.** 2009. Discrimination of viable and dead fecal Bacteroidales bacteria by quantitative PCR with propidium monoazide. Appl Environ Microbiol **75:**2940-2944.

19. **Kralik P, Nocker A, Pavlik I.** 2010. *Mycobacterium avium* subsp *paratuberculosis* viability determination using F57 quantitative PCR in combination with propidium monoazide treatment. Int J Food Microbiol **141:**S80-S86.

20. **Schloss PD, Gevers D, Westcott SL.** 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLOS One **6:**e27310.

21. **The R Development Core Team.** 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

22. **Viechtbauer W.** 2010. Conducting meta-analyses in R with the metafor package. J Stat Soft **36:**1-48.

23. **Magurran AE, Henderson PA.** 2003. Explaining the excess of rare species in natural species abundance distributions. Nature **422:**714-716.

24. **Douglas Bates MM, Ben Bolker, Steve Walker.** 2014. Fitting Linear Mixed-Effects Models using lme4. R package version 1.1-7.

25. **Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS, Wolfgang MC, Boucher R, Gilpin DF, McDowell A, Elborn JS.** 2008. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am J Resp Crit Care **177:**995-1001.

26. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Bruce KD.** 2004. Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. J Clin Microbiol **42:**5176-5183.

27. **Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM.** 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. ISME J **5:**20-29.

28. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, Connett GJ, Bruce KD.** 2006. Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. J Clin Microbiol **44:**2601-2604.

29. **Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu WH, Lakshmanan A, Wade WG.** 2010. The Human Oral Microbiome. J Bacteriol **192:**5002-5017.

30. **Berger WH, Parker FL.** 1970. Diversity of planktonic foraminifera in deep-sea sediments. Science **168:**1345-1347.

31. **Harris JK, De Groote MA, Sagel SD, Zemanick ET, Kapsner R, Penvari C, Kaess H, Deterding RR, Accurso FJ, Pace NR.** 2007. Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. PNAS USA **104:**20529-20533.

32. **van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD.** 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. ISME J **5:**780-791.

33. 33. **van der Gast CJ, Cuthbertson L, Rogers GB, Pope C, Marsh RL, Redding GJ, Bruce KD, Chang AB, Hoffman LR.** 2014. Three Clinically Distinct Chronic Pediatric Airway Infections Share a Common Core Microbiota. Ann Am Thora Soc.

34. **Mojon P, Budtz-Jorgensen E, Pazos E.** 2002. Attitude toward oral health among institutionalized elders. J Dent Res **81:**A178-A178.

35. **Gibson DJ, Baer SG, Klopf RP, Reed LK, Wodika BR, Willand JE.** 2013. Limited effects of dominant species population source on community composition during community assembly. J Veg Sci **24:**429-440.

36. **Lynch SV, Bruce KD.** 2013. The Cystic Fibrosis Airway Microbiome. Csh Perspect Med **3**.

37. **Zhao JC, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ.** 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. PNAS USA **109:**5809-5814.

38. **Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu WH, Lakshmanan A, Wade WG.** 2010. The Human Oral Microbiome. J Bacteriol **192**:5002-5017.

3.6 Appendix

Table A3.1 The raw sequence data reported in this Chapter have been deposited in the European Nucleotide Archive short read database

Study Accession Number: ERP007059, Sample Accession Number: RS551400. A list of barcodes used and their associated samples are listed below.

Barcode	Sample	PMA	Barcode	Sample	PMA	Barcode	Sample	PMA
ACAGTGCTTCAT	10	No	ACTTGTAGCAGC	29	Yes	AGTCACATCACT	47	No
ACATCACTTAGC	10	Yes	AAGCTGCAGTCG	3	No	AGTCTACTCTGA	47	Yes
ACATGATCGTTC	11	No	AATCGTGACTCG	3	Yes	AGTCCATAGCTG	48	No
ACATTCAGCGCA	11	Yes	ACTGTGACTTCA	30	No	AGTCTCGCATAT	48	Yes
ACATGTCACGTG	12	No	AGAACACGTCTC	30	Yes	ACACATGTCTAC	5	No
ACCACATACATC	12	Yes	AGACCGTCAGAC	31	No	ACACGGTGTCTA	5	Yes
ACCAGCGACTAG	14	No	AGACTGCGTACT	31	Yes	AGTGCGATGCGT	50	No
ACCTCGATCAGA	14	Yes	AGACGTGCACTG	32	No	AGTGTCACGGTG	50	Yes
ACCTGTCTCTCT	15	No	AGAGAGCAAGTG	32	Yes	AGTTAGTGCGTC	52	No
ACGAGTGCTATC	15	Yes	AGAGCAAGAGCA	33	No	AGTTCTACGTCA	52	Yes
ACGACGTCTTAG	16	No	AGAGTCCTGAGC	33	Yes	ATACACGTGGCG	54	No
ACGATGCGACCA	16	Yes	AGATCGGCTCGA	35	No	ATACGTCTTCGA	54	Yes
ACGCGATACTGG	18	No	AGATGTTCTGCT	35	Yes	ATACTATTGCGC	55	No
ACGCTATCTGGA	18	Yes	AGATCTCTGCAT	36	No	ATAGCTCCATAC	55	Yes
ACGCTCATGGAT	19	No	AGCACACCTACA	36	Yes	ATACTCACTCAG	56	No
ACGGTGAGTGTC	19	Yes	AGCACGAGCCTA	37	No	ATAGGCGATCTC	56	Yes
ACGGATCGTCAG	20	No	AGCAGTCGCGAT	37	Yes	ATATGCCAGTGC	58	No
ACGTACTCAGTG	20	Yes	AGCAGCACTTGT	38	No	ATCACTAGTCAC	58	Yes
ACGTCTGTAGCA	21	No	AGCATATGAGAG	38	Yes	ACACGAGCCACA	6	No
ACGTGCCGTAGA	21	Yes	AATCAGTCTCGT	4	No	ACACTAGATCCG	6	Yes
ACGTGAGAGAAT	22	No	ACACACTATGGC	4	Yes	ATGCACTGGCGA	67	No
ACGTTAGCACAC	22	Yes	AGCGACTGTGCA	40	No	ATGCCTGAGCAG	67	Yes
ACTACAGCCTAT	23	No	AGCGCTGATGTG	40	Yes	ATGGATACGCTC	69	No
ACTAGCTCCATA	23	Yes	AGCGTAGGTCGT	41	No	ATGGCGTGCACA	69	Yes
ACTACGTGTGGT	24	No	AGCTCCATACAG	41	Yes	ACACTGTTCATG	7	No
ACTATTGTCACG	24	Yes	AGCTATCCACGA	42	No	ACAGAGTCGGCT	7	Yes
ACTCACGGTATG	25	No	AGCTCTCAGAGG	42	Yes	ATTATCGTGCAC	73	No
ACTCGATTCGAT	25	Yes	AGCTGACTAGTC	43	No	CAACACGCACGA	73	Yes
ACTCAGATACTC	26	No	AGGACGCACTGT	43	Yes	GTCGTGTGTCAA	77	No
ACTCGCACAGGA	26	Yes	AGGTGTGATCGC	45	No	GTCTATCGGAGT	77	Yes
ACTGACAGCCAT	28	No	AGTACTGCAGGC	45	Yes	GTCTACACACAT	78	No
ACTGTACGCGTA	28	Yes	AGTACGCTCGAG	46	No	GTCTCATGTAGG	78	Yes
ACTGTCGAAGCT	29	No	AGTAGTATCCTC	46	Yes			

Table A3.2 Species-level identities of detected bacterial taxa identified from 36 sputum samples collected from CF patients.

Given the length of the ribosomal sequences analysed (approx 350bp), these identities should be considered putative. Ae denotes aerobe and An, Anaerobe. Only strict anaerobes were classified as anaerobes, whereas aerobes, facultative anaerobes, and microaerophiles were classified as aerobes. O indicates members of the oral microbiota according to the Human Oral Microbiome database (38). Organisms considered to be common are highlighted in bold.

Class	Family	Species	Anaerobes	Oral
Actinobacteria	Actinomycetaceae	*Actinobaculum massiliense*	Ae	
		Actinomyces massiliensis	Ae	O
		Actinomyces odontolyticus	Ae	O
		Actinomyces oris	Ae	O
	Corynebacteriaceae	*Corynebacterium durum*	Ae	O
	Corynebacteriaceae	*Corynebacterium pseudodiphtheriticum*	Ae	
	Microbacteriaceae	*Microbacterium oxydans*	Ae	
	Micrococcaceae	*Rothia dentocariosa*	Ae	
	Nocardiaceae	*Nocardia cyriacigeorgica*	Ae	
	Propionibacteriaceae	*Propionibacterium acnes*	Ae	O
		Propionibacterium propionicum	Ae	O
	Bifidobacteriaceae	*Scardovia wiggsiae*	An	
	Coriobacteriaceae	*Atopobium parvulum*	An	O
		Cryptobacterium curtum	An	O
		Olsenella uli	An	O
		Slackia exigua	An	O
Bacteroidia	Bacteroideceae	*Bacteroides acidifaciens*	An	
	Porphyromonadaceae	*Barnesiella intestinihominis*	An	
		Paludibacter propionicigenes	An	
		Porphyromonas catoniae	An	
		Porphyromonas endodontalis	An	O
		Tannerella forsythia	An	O
	Prevotellaceae	*Alloprevotella rava*	An	
		Prevotella bivia	An	O
		Prevotella denticola	An	O
		Prevotella loescheii	An	O
		Prevotella maculosa	An	O

Table A3.2 Continued

Class	Family	Species	Anaerobes	Oral
		Prevotella melaninogenica	**An**	**O**
		Prevotella nanceiensis	An	
		Prevotella nigrescens	An	O
		Prevotella oris	An	O
		Prevotella pallens	An	O
		Prevotella pleuritidis	An	O
		Prevotella shahii	An	O
		Prevotella tannerae	An	O
		Prevotella timonensis	An	
Flavobacteria	Flavobacteriaceae	*Capnocytophaga gingivalis*	Ae	O
		Capnocytophaga haemolytica	Ae	O
		Capnocytophaga ochracea	Ae	O
		Capnocytophaga sputigena	Ae	O
Bacilli	Bacillales Incertae Sedis XI	*Gemella morbillorum*	Ae	O
	Staphylococcaceae	*staphylococcus aureus*	Ae	O
	Aerococcaceae	*Abiotrophia defectiva*	Ae	O
	Carnobacteriaceae	*Granulicatella adiacens*	Ae	O
	Lactobacillaceae	*Lactobacillus fermentum*	Ae	O
		Lactobacillus johnsonii	Ae	O
		Lactobacillus salivarius	Ae	O
	Streptococcaceae	***Streptococcus parasanguinis***	**Ae**	**O**
		Streptococcus pneumoniae	**Ae**	**O**
		Streptococcus sanguinis	Ae	O
Clostridia	Clostridiaceae	*Clostridium bolteae*	An	
	Clostridiales Incertae Sedis XI	*Parvimonas micra*	An	O
		Peptoniphilus lacrimalis	An	O
		Mogibacterium diversum	An	O
	Eubacteriaceae	*Eubacterium brachy*	An	O
		Eubacterium infirmum	An	O
	Lachnospiraceae	*Butyrivibrio hungatei*	An	O
		Catonella morbi	An	O
		Howardella ureilytica	An	
		Oribacterium sinus	An	O
		Shuttleworthia satelles	An	O
		Stomatobaculum longum	An	
		Lachnoanaerobaculum orale	An	O
	Peptococcaceae	*Peptococcus niger*	An	
	Peptostreptococcaceae	*Peptostreptococcus stomatis*	An	O

Table A3.2 Continued

Class	Family	Species	Anaerobes	Oral
	Veillonellaceae	*Anaeroglobus geminatus*	An	O
		Dialister invisus	An	O
		Dialister micraerophilus	An	O
		Dialister pneumosintes	An	O
		Megasphaera micronuciformis	An	O
		schwartzia succinivorans	An	O
		Selenomonas noxia	An	O
		Selenomonas sputigena	An	O
		Veillonella dispar	**An**	**O**
		Veillonella ratti	An	O
Fusobacteria	Fusobacteriaceae	*Fusobacterium nucleatum*	An	O
	Leptotrichiaceae	*Leptotrichia buccalis*	An	O
		Leptotrichia wadei	An	O
		Sneathia sanguinegens	An	O
Alphaproteobacteria	Sphingomonadaceae	*Sphingomonas leidyi*	Ae	
Betaproteobacteria	Alcaligenaceae	*Achromobacter xylosoxidans*	Ae	O
	Burkholderiaceae	*Lautropia mirabilis*	Ae	O
		Ralstonia pickettii	Ae	
	Comamonadaceae	*Comamonas testosteroni*	Ae	
	Neisseriaceae	*Kingella oralis*	Ae	O
		Neisseria bacilliformis	Ae	O
		Neisseria flavescens	Ae	O
		Neisseria oralis	Ae	O
Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*	Ae	O
Gammaproteobacteria	Cardiobacteriaceae	*Cardiobacterium valvarum*	Ae	
	Enterobacteriaceae	*Escherichia coli*	Ae	O
	Pasteurellaceae	*Haemophilus haemolyticus*	Ae	
		Haemophilus parainfluenzae	Ae	O
	Moraxellaceae	*Moraxella osloensis*	Ae	O
	Pseudomonadaceae	*Pseudomonas putida*	Ae	
		Pseudomonas aeruginosa	**Ae**	**O**
	Xanthomonadaceae	*Stenotrophomonas maltophilia*	Ae	O
Mollicutes	Mycoplasmataceae	*Mycoplasma salivarium*	Ae	O

Chapter 4: Collection to storage time affects respiratory samples

Time between sputum sample collection and storage significantly influences bacterial sequence composition from Cystic Fibrosis respiratory infections

Cuthbertson L, Rogers G.B., Walker A.W., Oliver A., Hafiz T., Hoffman L.R., Carroll M.P., Parkhill J., Bruce K.D., van der Gast C.J. (2014) **Bacterial respiratory infection composition changes with increasing time between sputum sample collection and storage**. *Journal of Clinical Microbiology*, **52**(8):3011-3016

As outlined in the previous chapter, accurate investigation of the microbial community within CF sputum is hugely important to inform treatment choices. The use of PMA allows only the viable community to be analysed preventing free DNA and DNA from dead or damaged cells from being included in analysis, therefore giving a true picture of the community within the lung. As a continuation of this, Chapter 4 investigates how sample handling can affect the bacterial community within sputum. This Chapter has been published in the Journal of Clinical Microbiology.

Main title: Time between sputum sample collection and storage significantly influences bacterial sequence composition from Cystic Fibrosis respiratory infections

Running title: Collection to storage time affects respiratory samples

Leah Cuthbertson[1,2], Geraint B. Rogers[2,7], Alan W. Walker[3], Anna Oliver[1], Tarana Hafiz[4], Lucas R. Hoffman[5,6], Mary P. Carroll[4], Julian Parkhill[3], Kenneth D. Bruce[2], Christopher J. van der Gast[1,a]

[1] *NERC Centre for Ecology & Hydrology, Wallingford, OX10 8BB, UK*

[2] *Institute of Pharmaceutical Science, Molecular Microbiology Research Laboratory, King's College London, London, SE1 9NH, UK*

[3] *Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK*

[4] *Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton, SO16 6YD, UK*

Departments of [5]Pediatrics and [6]Microbiology, University of Washington, Box 356320, HSB K328A, Seattle, WA 98105, USA

[7] SAHMRI Infection and Immunity Theme, School of Medicine, Flinders University, Bedford Park, Adelaide, Australia

[a] To whom correspondence should be addressed: E-mail: cjvdg@ceh.ac.uk

The authors declare no conflict of interest.

Data deposition: The raw sequence data reported in this paper have been deposited in the NCBI Short Read Archive database (Accession number. SRP036061). A table listing the barcodes used is included as Supplementary Table 2.

4.1 Abstract

Spontaneously expectorated sputum is traditionally used as the sampling method for the investigation of lower airway infections. Whilst guidelines exist for the handling of these samples for culture-based diagnostic microbiology, there is no comparable consensus on their handling prior to culture-independent analysis. The increasing incorporation of culture-independent approaches in diagnostic microbiology means it is of critical importance to assess potential biases. The aim of this study was to assess the impact of delayed freezing on culture-independent microbiological analyses, and to identify acceptable parameters for sample handling. Sputum samples from eight adult cystic fibrosis (CF) patients were collected and aliquoted into sterile Bijou bottles. Aliquots were stored at room temperature before freezing at -80°C for increasing intervals up to 72 hour period. Samples were treated with propidium monoazide, to distinguish live from dead cells, prior to DNA extraction, and 16S rRNA gene pyrosequencing was used to characterise the bacterial composition. Substantial variation was observed in samples with high diversity bacterial communities over time, whereas low diversity communities dominated by recognised CF pathogens varied little regardless of time to freezing. Partitioning into common and rare species demonstrated that the rare species drove changes in similarity. The percentage abundance of anaerobes over the study significantly decreased after 12 hours at room temperature (P=0.008). Failure to stabilise samples at -80°C within 12 hours of collection results in significant changes in the detected community composition.

4.2 Introduction

Next generation sequencing techniques are increasingly being used to characterise respiratory microbiota in many lung diseases, including cystic fibrosis (CF) lower airway microbiota (eg.1-4). These analyses have revealed microbial communities within the CF lung to be more complex and diverse than previously considered. Importantly, they have also detected many bacterial species that would not be reported by standard diagnostic microbiology (eg. 5,6), as well as identifying relationships between microbiota characteristics and host age, lung function, and disease progression (7-9).

In the majority of cases, these investigations relied on spontaneously expectorated sputum as a means to sample the bacterial communities in the lower airways. Sputum is favoured due to ease of collection, and the fact that culture-based microbiological studies, of adult patients have traditionally used sputum samples as a basis for microbiological analysis. Whilst guidelines exist for the handling of respiratory samples for culture-based diagnostic microbiology (10), there is no consensus on how such samples should be handled to ensure that resulting culture-independent analyses yield results reflecting the microbes therein. With the increasing move towards the incorporation of culture-independent methods into diagnostic microbiology (11) it is increasingly important to identify and minimise areas of potential bias

Post-collection sample transportation and storage represents a period during which changes can occur in bacterial community of clinical samples, resulting in analytical bias due to, for example, bacterial proliferation, cell death or degradation of nucleic acids. In order to minimise these biases sputum samples collected for culture-independent analyses are typically stored at -80°C. However, many clinical sites, including those that treat CF patients, lack ready access to ultra-low temperature freezers; the standard recognised method for maintenance of sample integrity and biobanking. As a result, sputum samples may remain at room temperature for extended periods, impacting both traditional and culture-independent analyses.

A prior study used 16S rRNA gene pyrosequencing of a single sample to assess the effect of extended periods of incubation at room temperature on bacterial community profiles, but did not find significant divergence in community composition over the study period (12). Conversely, in an earlier study, using ribosomal transcripts to examine the V3 region of the 16S rRNA gene by quantitative PCR and denaturing gradient gel electrophoresis (DGGE), significant divergence in bacterial quantitation and community profiling was observed (13). RNA-based approaches have the advantage of limiting analysis to active cells. However, a related exclusion of non-viable cells and extracellular DNA can be achieved in DNA-based analysis through the treatment of samples with propidium monoazide (PMA), as we have demonstrated in previous analyses of CF sputum (14-16).

We hypothesised that the period of time between sample collection and stabilisation by freezing is significantly related to the resultant bacterial community composition, as determined by 16S rRNA gene pyrosequencing in combination with PMA treatment. From this our overarching aim was to determine an appropriate window of time from sample collection to storage at -80°C that would allow reliable culture-independent microbiological analysis of sputum samples.

4.3 Methods

4.3.1 Sample collection

Sputum samples were collected, under full ethical approval from the Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26), from eight adult patients attending the regional Cystic Fibrosis Centre in Southampton General Hospital for treatment for clinical exacerbation. Patients were chosen for their abilities to provide sputum samples of 3 ml or more.

Samples were collected during physiotherapy and immediately aliquoted into sterile 5 ml Bijou bottles and stored at room temperature until freezing to -80°C at specified intervals. Samples at t=0 were stored at -80°C immediately. The remaining samples were held at room temperature before storage at -80°C, for 1, 3, 6, 9, 12, 18, 24, 36, 48, 60 and 72 hours. The 72 hour storage period was chosen to allow investigation of changes in the bacterial community beyond the maximum 48 hour storage recommended by Health Protection England for culture-based diagnostic microbiology (10).

4.3.2 DNA extraction and Pyrosequencing

Sputum samples were washed three times with 1x phosphate buffered saline. Free DNA and DNA from non-viable cells were excluded from analysis via crosslinking with PMA (14, 15) prior to DNA extraction, as previously described (16). Bacterial Golay Barcode encoded FLX amplicon pyrosequencing was performed using the primer 338F (5'- ACTCCTACGGGAGGCAGCAG) and 926R (5'- CCGTCAATTCMTTTRAGT). Initial generation 16S rRNA gene amplicons involved a one step PCR of 25 cycles using AccuPrime™ Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA). 454 pyrosequencing using the Lib-L kit was performed at the Wellcome Trust Sanger Institute, Hinxton, UK.

4.3.3 Sequence analysis

The mothur sequencing analysis platform was used to analyse the resulting data (17). Failed sequence reads, low quality sequence ends, tags and primers, were initially removed, followed by sequences below 400bp and any sequences that included ambiguous base calls and homopolymers longer than 8 bases. Chimeras were removed in mother using the Perseus software program (18). Sequences were assembled into operational taxonomic units (OTUs) clusters at 97% identity, to give an approximation of species (19), and identified using the RDP reference database. Representative sequences were used to give an appropriate species level identification for OTUs using NCBI BlastN. The raw sequence data

generated within the current study have been submitted to the NCBI Short Read Archive database under the study accession number SRP036061, Table S2.

4.3.4 Statistical analysis

All statistical analysis was performed in R (20). Three complementary measures of diversity were used as previously described (9) to identify changes in bacterial diversity in the same sample kept at room temperature for different durations; species richness (*S**), Shannon-Wiener (*H*'), and Simpson's (1-*D*) indices of diversity (21). The Bray-Curtis similarity index (21) was calculated using randomised re-sampling to compare changes in community composition over time at room temperature. To avoid potential biases due to sampling depth, randomised re-sampling with a uniform resample size (n=200 to match smallest sample size) was carried out (22), as described previously (9, 23). 1000 iterations of each resampling were performed to obtain the mean diversity and similarity coefficients and standard deviation of the mean. The Berger-Parker (*d*) measure of dominance was calculated using the BiodiversityR package (24).

Bacterial species at t=0 for each patient were partitioned into common and rare species using the inflection point method from rank abundance curves as previously described (25). A one-way ANOVA was calculated with two independent categorical variables, time and partition (common or rare), to compare the difference in similarity between common and rare species (20). The post-hoc Tukey honest significant difference (HSD) test was used in conjunction with the ANOVA to compare treatment means in order to find significant differences (20).

Change in anaerobe abundance over time was investigated using nlme (26) to fit mixed effect models, r^2 values were calculated using the MuMIn package (27).

4.4 Results

Sputum samples were collected from eight patients and then aliquoted into 12 equal portions and stored at room temperature (mean ± standard error of mean (SE), 20.1°C±0.1) for specified intervals over a 72hour study period. Of the 96 sample aliquots sequenced, 12 were excluded from further analysis due to insufficient number of sequences (i.e. fewer than 200 sequences). A total of 182,989 bacterial sequences (n=84, 2178±250SE/sample) were generated from 84 samples, identifying 51 genera and 78 distinct OTUs classified to species level (Table S1).

4.4.1 Bacterial diversity

Changes in bacterial diversity were assessed over the study period using the recognised measures of diversity, species richness (*S**, the total number of species), Shannon-Wiener index (*H*', a metric accounting for both number and relative abundance of species) and Simpson's diversity index (1-*D*, a measure of the probability that two species randomly selected from a sample will differ). S*, H' and 1-D were calculated for each sample from each patient over the 72 hour study period, using randomised re-sampling as previously described (23), as pairwise comparisons are affected by large differences in sample size (n) (28).

High levels of variation were observed when examining diversity measures, both between and within patients. In order to investigate this variation, sample diversity at t=0 was plotted for S*, H' and 1-D (Fig. 1).

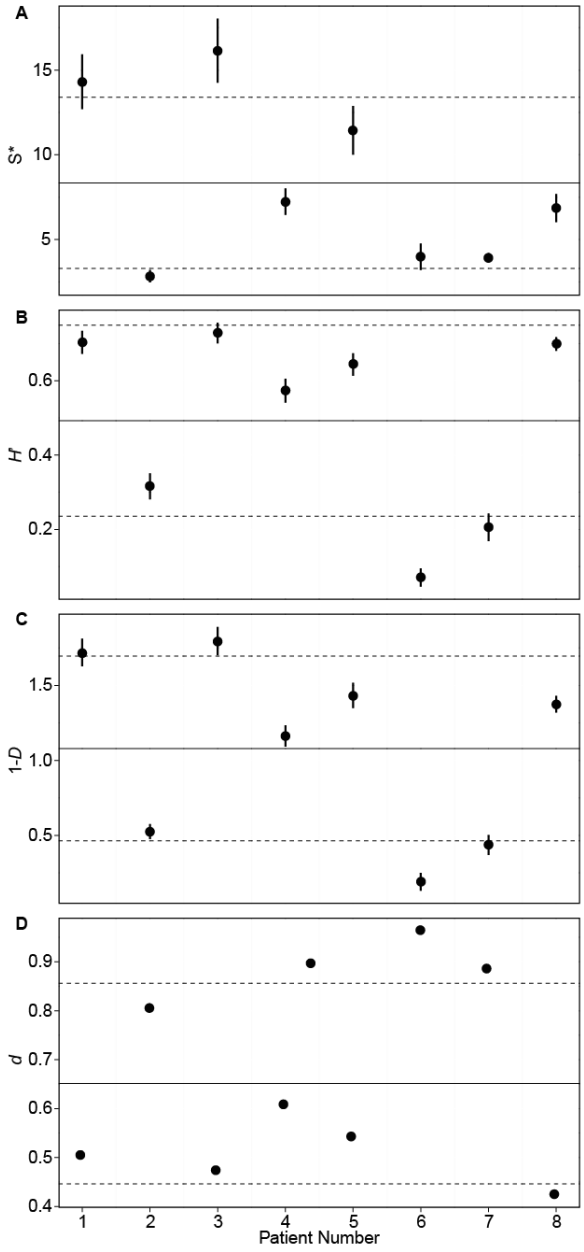


Fig 4.1 Comparison of diversity and dominance of bacterial communities across patients at *t*=0. Values of (a) species richness (*S**), (b) Shannon-Wiener index of diversity (*H*'), and (c) Simpson's index of diversity (1-*D*) as shown. The three diversity indices were calculated with a uniform re-sample size following 1000 iterations in each instance. Error bars represent the standard deviation of the mean (*n* = 1000). In each instance, the overall mean (solid line)

and the standard deviation of the mean (dashed lines) across patients are shown. Also given is (d) Berger-Parker index of dominance (*d*).

Typically, low diversity communities are dominated by a few highly abundant species whereas high diversity communities are characterised by species that are more evenly distributed in their abundances (29). In order to explain the observed variation in diversity in the current study, the Berger Parker index of dominance (*d*, the proportional abundance of the most abundant organism) was calculated for all t=0 samples (Fig. 1) and for samples at all time points (Fig. S1). Here, we also observed that low diversity (as defined by *S*, H'* and 1-*D*) was related to high dominance (*d*) in sample communities: for example, the t=0 community of patient 6 was dominated by *Pseudomonas aeruginosa* (S*= 4, *d*= 0.96), and the t=0 communities of Patients 2 and 7 were dominated by *Achromobacter xylosoxidans* (Patient 2; S*= 3, *d*= 0.88, Patient 7; S*= 4, *d*= 0.80), while Patient 1 showed a much more diverse t=0 community (S*= 14, *d*= 0.51), Figs. 1 and S1. Given the high level of variation in diversity between patients, linked to species dominance, these measures were unsuitable as metrics to indicate changes over time, suggesting that comparisons of community similarity over time would be more appropriate.

4.4.2 Bacterial community membership

The Bray-Curtis measure of similarity (S_{BC}, which accounts for the number and abundance of species present in each community and those that are shared) was used to compare changes in community composition between samples, resulting in a value between 0 and 1 (higher values indicating greater similarity). As with the diversity measures, community composition were compared between the sample at t=0 and each subsequent sample. Using randomised re-sampling, total change in S_{BC} similarity was assessed (22).

For both PCR and sequencing, sampling bias can result in variation between repeat samples of the same community (12). In order to evaluate whether changes in similarity across the sampling period were due to true community changes or within sample variation, a cut off

value for similarity was calculated using eight samples each sequenced in triplicate. The overall mean S_{BC} similarity between sample replicates was 0.782±0.1(SE) (n=24), therefore similarity values, below 0.682 (0.782-0.10SE), between t=0 and subsequent samples were judged to be different from the original sample.

The mean change in similarity over the study period was not judged to be significant, when accounting for within sample variation (Fig. 2(A)). When diversity measures were examined for individual patients, high levels of variation were observed (Fig. 3). These results were compared to the value of dominance calculated previously using *d* (Fig. 1). The greater the relative abundance of the dominant species, the lower the variation in community similarity over the study period (P=0.03, r^2=0.05). This finding indicated that samples more highly dominated by a single species, and hence having low overall diversity, were less likely to show a significant change in community similarity with longer time at room temperature prior to freezing.

To further examine how dominant species affect community similarity, rank abundance curves were used to partition sample communities at t=0 into common and rare species groups (Fig. S2)(25). Change in S_{BC} similarity was calculated, from t=0, for the partitioned groups revealing much greater variation in species characterised as rare Fig 2. Using ANOVA a significant difference in similarity was observed between the common and rare species ($F_{(1,148)}$=77.93, P<0.001). Post-hoc Tukey HSD testing revealed that the difference in similarity between t=0 and the rare species were on average 30.93% lower than the differences observed in the common species.

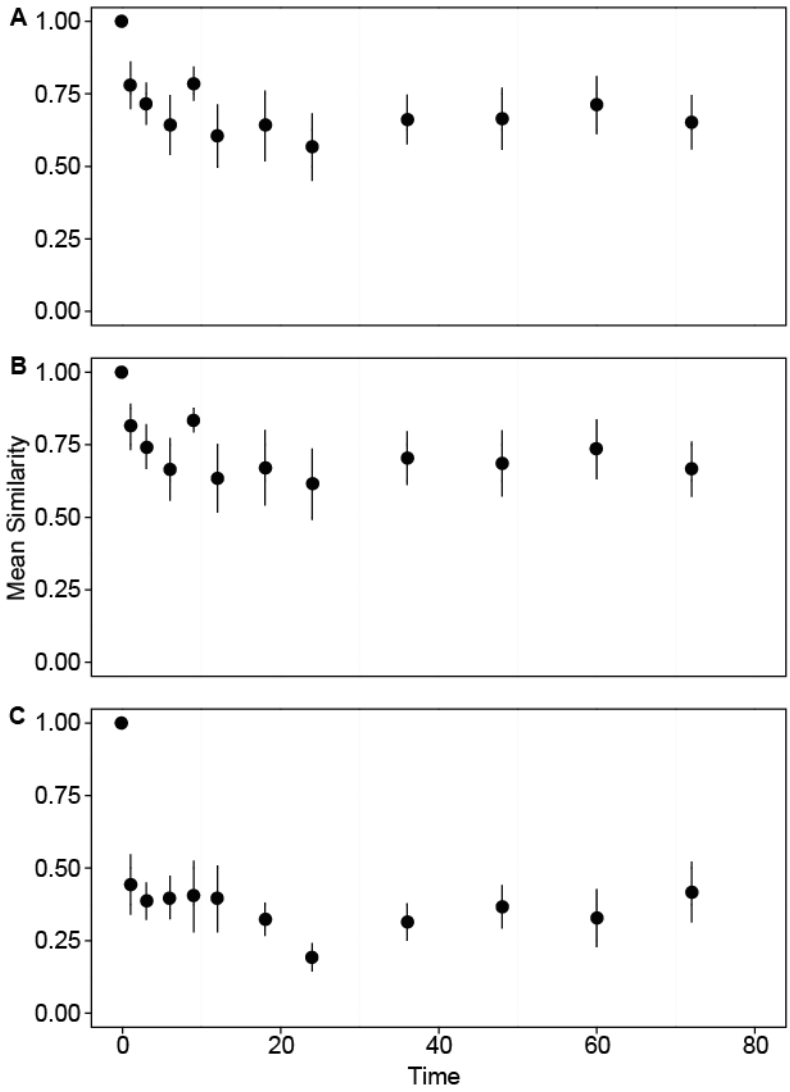


Fig. 4.2 Mean changes in bacterial community composition, from t=0 across patients over time using the Bray-Curtis index of similarity for (a) whole communities, (b) common and (c) rare species groups. Error bars represent the standard deviation of the mean (*n* = 8).

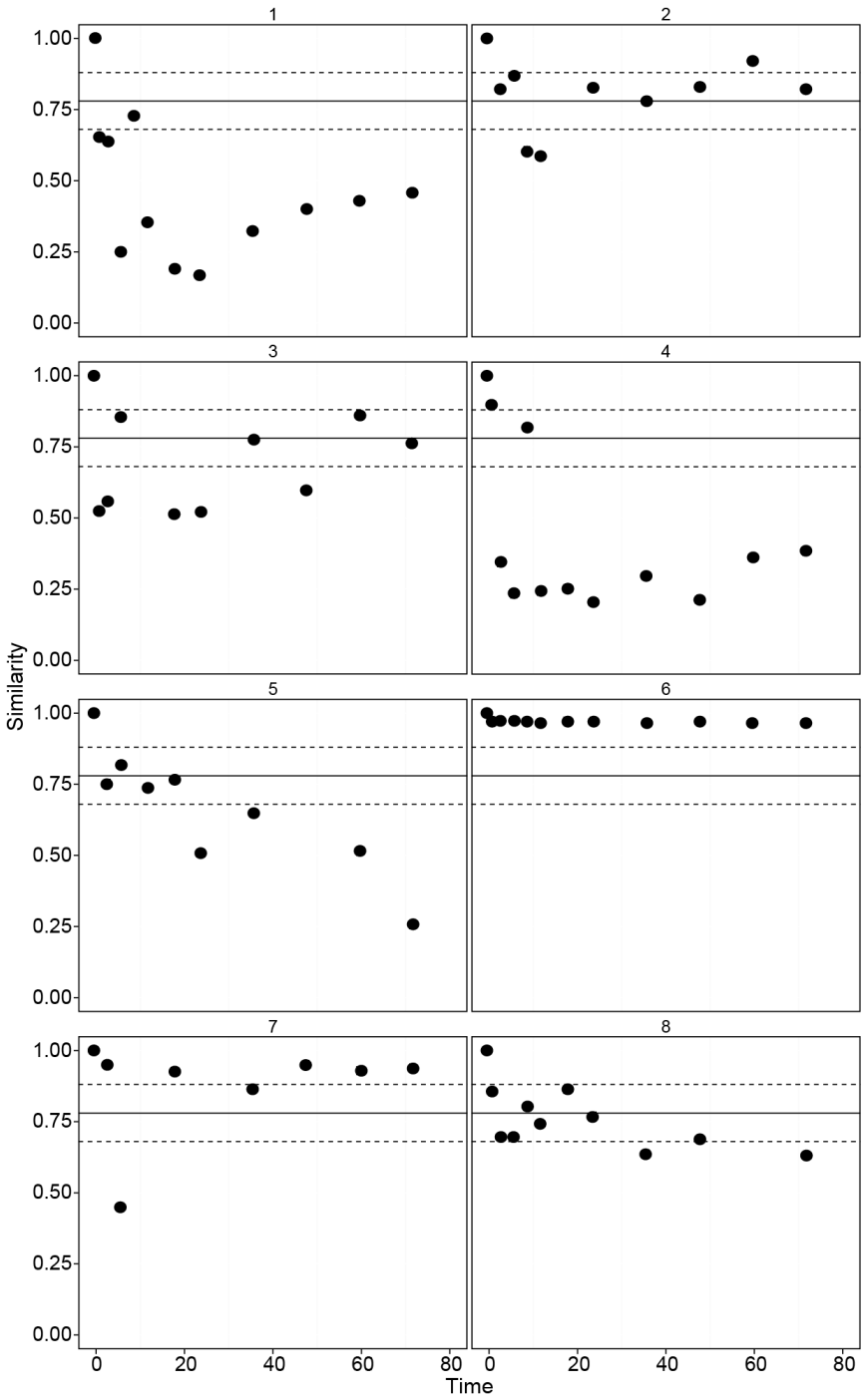


Fig. 4.3 Changes in community composition from t=0, for each patient over time, using the Bray-Curtis index of similarity. Solid lines represent the overall mean similarity from within sample replicates and dashed lines represent the standard error of the mean (*n* = 24).

4.4.3 Differential impact on aerobic and anaerobic species

Over the study period, samples were aliquoted and stored in sterile sample containers, it was expected that this would result in a decrease in anaerobic species due to prolonged exposure to atmospheric oxygen resulting in preferential conditions for aerobic populations. To that end, bacteria were partitioned into aerobic/facultative anaerobes and strict anaerobic species. A mixed effects model was used to investigate the change in percentage abundance of anaerobic species present in each sample over the study period. The best-fit distribution was a second order polynomial relationship (r^2=0.08, P=0.004) (Fig. 4). This distribution indicated a consistent decline in the relative abundance of anaerobic species over the first 48 hours, followed by an increase during the following 24 hours. Using mixed effect models, the decline in anaerobes was found to represent a significant divergence from the original sample after 18 hours (P=0.008). This decline continued until after 48 hours of storage at room temperature, at which point the percentage abundance of anaerobes started to increase. This finding suggests that changes in anaerobe abundance was due to sputum storage in sealed containers allowing anaerobic organisms to proliferate after available oxygen had been depleted, resulting in community divergence.

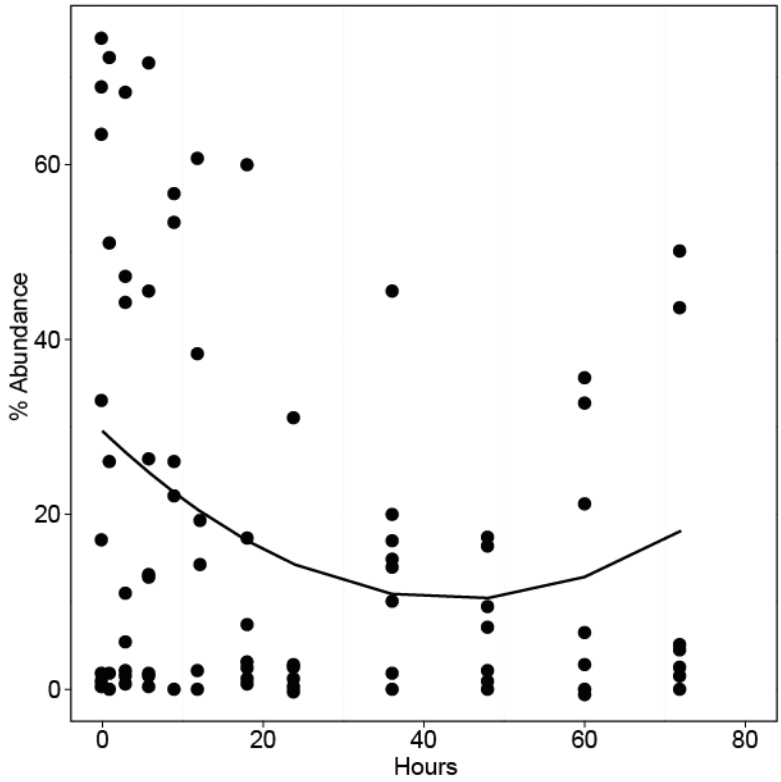


Fig. 4.4 Changes in anaerobic species abundance for all patients over time. Solid circles represent percentage abundance of anaerobic species for each patient at a given time point. A second order polynomial model has been fitted to the data (r^2 = 0.08, P = 0.004).

4.5 Discussion

The handling and storage of respiratory samples can substantially bias the results of microbiological analyses. The aim of the current study was to determine an acceptable period of time between sputum sample collection and storage prior to using next generation sequencing to characterise CF lower airway microbiota. Previous studies based on diagnostic bacterial culture suggest that 48 hours can elapse from specimen collection to either freezing or processing and still provide comparable results (10). However, with the increasing incorporation of culture-independent approaches to diagnostic microbiology (11) it is vital to identify and mitigate all potential biases relevant for these more sensitive, culture-independent techniques.

Changes in bacterial diversity, as a result of storage at room temperature for different intervals prior to freezing, were assessed within samples from each patient. We found that diversity was highly variable across patients at t=0 (Fig. 1) and within patients over storage time, post collection (Fig. S1). This variability could be attributed to differences in dominance within sample communities. Freshly collected (t=0) samples with low species diversity were found to be highly dominated by recognised CF pathogens; conversely, those samples with a more diverse communities were not dominated by a particular bacterial species. Furthermore, communities dominated by fewer species changed less with increasing time at room temperature compared to those with higher diversities. Given the differences in diversity between and within patient samples, commonly used measures of diversity, e.g. species richness, Shannon-Wiener diversity index and Simpson's diversity index, were unsuitable as a metric of change in the bacterial communities during storage.

Using the Bray-Curtis index of similarity, we assessed how community composition changed from the original (t=0) sample. As previous studies have indicated that bacterial communities within the CF lungs are not homogenously distributed (30), this may result in a variation within different portions of a single sample (12). In order to account for within sample variance, eight CF sputum samples were sequenced in triplicate, using different sample

aliquots for each replicate, and the similarity between replicates was calculated. Then, we analysed all samples for all subjects, finding that the shortest period of sample storage within which a change in similarity was observed, beyond that expected for within-sample variation, was 1 hour after sputum collection (Fig. 2). In addition where a community was dominated by few or one species, the variation in community similarity was found to be significantly lower than in more diverse communities. However, since it is difficult to guess the microbial diversity in a sputum sample *a priori*, our results suggest that sputum samples should be frozen within 1 hour of collection in order to obtain the best possible representation of the true community using culture independent analyses.

Previous studies have demonstrated the value of partitioning bacterial communities in respiratory infections into common and rare species groups (9). Categorization of component species provides useful insights into communities that would be neglected without such a distinction (3, 9). When samples were partitioned into common and rare species in the current study, a greater level of community stability with difference in storage, was associated with species defined as common compared to rare species, suggesting that characterisation of the latter group will be most affected by a delay in sample freezing. This effect accounted for the greater change in similarity observed in more diverse communities that have a wider range of rare species present within the community (Fig. 2).

We hypothesised that prolonged exposure to atmospheric levels of oxygen would result in a decrease in the relative abundance of viable, strictly anaerobic species within the sputum samples. Despite high variability in anaerobe relative abundance between samples, a statistically significant 2^{nd} order polynomial relationship was found between storage duration at room temperature and anaerobe abundance, with the latter decreasing for up to 48 hours, followed by an increase after that time (Fig. 4). A potential explanation for this relationship is the reduction in oxygen tension as a result of growth of aerobic and aerotolerant species in the sealed Bijou containers. Furthermore, the decline in anaerobe abundance, which represents a major shift in the community, was found to be significantly different from t=0 after 12 hours (Fig. 4). This effect could potentially lead to under- or over-estimation of the

importance of anaerobic species in disease progression depending on the period elapsed between storage and freezing.

Sputum samples are one of the most widely used ways of sampling lower respiratory tract infections. With the moves toward incorporating culture independent techniques to analyse the microbial determinant of these conditions and make informed treatment choices. In the current study, we found that the optimal window for sample storage at room temperature before freezing at -80°C is within 1 hour of collection. In practical terms it may not be possible to store a sample within 1 hour of collection. In this event, our results indicate an acceptable window of up to 12 hours without significant divergence in community composition. Whilst this work has focused on CF airway infections, these findings are important for the analysis of microbiota from other respiratory conditions.

4.6 Acknowledgements

This study was supported by the UK Natural Environment Research Council (NE/H019456/1) to CJvdG, by the Wellcome Trust (WT 098051) to AWW and JP, and by the US National Institute of Health (K02HL105543 and P30 DK089507) to LRH.

4.7 References

1. **Tunney MM, Field TR, Moriarty TF, Patrick S, Doering G, Muhlebach MS, Wolfgang MC, Boucher R, Gilpin DF, McDowell A, Elborn JS**. 2008. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. Am J Resp Crit Care 177:995-1001.

2. **Zhao JC, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ.** 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. P Natl Acad Sci USA 109:5809-5814.

3. **Rogers GB, van der Gast CJ, Cuthbertson L, Thomson SK, Bruce KD, Martin ML, Serisier DJ**. 2013. Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. Thorax 68:731-737.

4. **Price K, Hampton T, Gifford A, Dolben E, Hogan D, Morrison H, Sogin M, O'Toole G**. 2013. Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. Microbiome 1:27.

5. **Rogers GB, Daniels TW, Tuck A, Carroll MP, Connett GJ, David GJ, Bruce KD**. 2009. Studying bacteria in respiratory specimens by using conventional and molecular microbiological approaches. BMC pulmonary medicine 9:14.

6. **Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC, Surette MG**. 2008. A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. P Natl Acad Sci USA 105:15070-15075.

7. **Cox MJ, Allgaier M, Taylor B, Baek MS, Huang YJ, Daly RA, Karaoz U, Andersen GL, Brown R, Fujimura KE, Wu B, Tran D, Koff J, Kleinhenz ME, Nielson D, Brodie EL, Lynch SV.** 2010. Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients. Plos One 5.

8. **Klepac-Ceraj V, Lemon KP, Martin TR, Allgaier M, Kembel SW, Knapp AA, Lory S, Brodie EL, Lynch SV, Bohannan BJM, Green JL, Maurer BA, Kolter R.** 2010. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. Environ Microbiol 12:1293-1303.

9. **van der Gast C, Walker A, Stressmann F, Rogers G, Scott P, Daniels T, Carroll M, Parkhill J, Bruce K**. 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. Isme J 5:780 - 791.

10. **Health Protection Agency**. 2012. Investigation of Bronchoalveolar Lavage, Sputum and Associated Specimens. UK Standards for Microbiology Investigations:http://www.hpa.org.uk/SMI/pdf.

11. **Pattison SH, Rogers GB, Crockard M, Elborn JS, Tunney MM.** 2013. Molecular detection of CF lung pathogens: Current status and future potential. Journal of Cystic Fibrosis 12:194-205.

12. **Zhao JC, Li J, Schloss PD, Kalikin LM, Raymond TA, Petrosino JF, Young VB, LiPuma JJ**. 2011. Effect of Sample Storage Conditions on Culture-Independent

Bacterial Community Measures in Cystic Fibrosis Sputum Specimens. Journal of Clinical Microbiology 49:3717-3718.

13. **Nelson A, De Soyza A, Bourke SJ, Perry JD, Cummings SP.** 2010. Assessment of sample handling practices on microbial activity in sputum samples from patients with cystic fibrosis. Lett Appl Microbiol 51:272-277.

14. **Nocker A, Sossa-Fernandez P, Burr MD, Camper AK.** 2007. Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol 73:5111-5117.

15. **Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD**. 2008. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diagn Microbiol Infect Dis 62:133-141.

16. **Rogers GB, Marsh P, Stressmann AF, Allen CE, Daniels TV, Carroll MP, Bruce KD.** 2010. The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. Clin Microbiol Infect 16:1656-1658.

17. **Schloss PD, Gevers D, Westcott SL.** 2011. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. Plos One 6.

18. **Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ**. 2011. Removing noise from pyrosequenced amplicons. Bmc Bioinformatics 12:38.

19. **Schloss PD, Handelsman J**. 2006. Toward a census of bacteria in soil. Plos Comput Biol 2:786-793.

20. **The R Development Core Team**. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

21. **Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H.** 2013. vegan: Community Ecology Package. R package version 2.0-7.

22. **Solow AR.** 1993. A Simple Test for Change in Community Structure. Journal of Animal Ecology 62:191-193.

23. **Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, Lilley AK, Oliver A, Carroll MP, Bruce KD, van der Gast CJ**. 2013. Reducing bias in bacterial community analysis of lower respiratory infections. Isme J 7:697-706.

24. **Kindt R, Coe R**. 2005. Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi:ISBN 978-992-9059-9222-9058.

25. **Siqueira T, Bini LM, Roque FO, Couceiro SRM, Trivinho-Strixino S, Cottenie K**. 2012. Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. Ecography 35:183-192.

26. **Pinheiro J BD, DebRoy S, Sarkar D, and the R Development Core Team**. 2013. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-110.

27. **Barton K**. 2013. MuMIn: Multi-model inference. R package version1.9.5.

28. **Gihring TM, Green SJ, Schadt CW**. 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. Environ Microbiol 14:285-290.

29. **Ager D, Evans S, Li H, Lilley AK, van der Gast CJ**. 2010. Anthropogenic disturbance affects the structure of bacterial communities. Environ Microbiol 12:670-678.

30. **Goddard AF, Dowd SE, Staudinger BJ, Wolcott RD, Aitken ML, Fligner CL, Singh PK.** 2010. Analysis of Microbial Diversity in Cystic Fibrosis Lung Explants Using Pyrosequencing. Pediatr Pulm:357-357.

4.8 Supplementary information

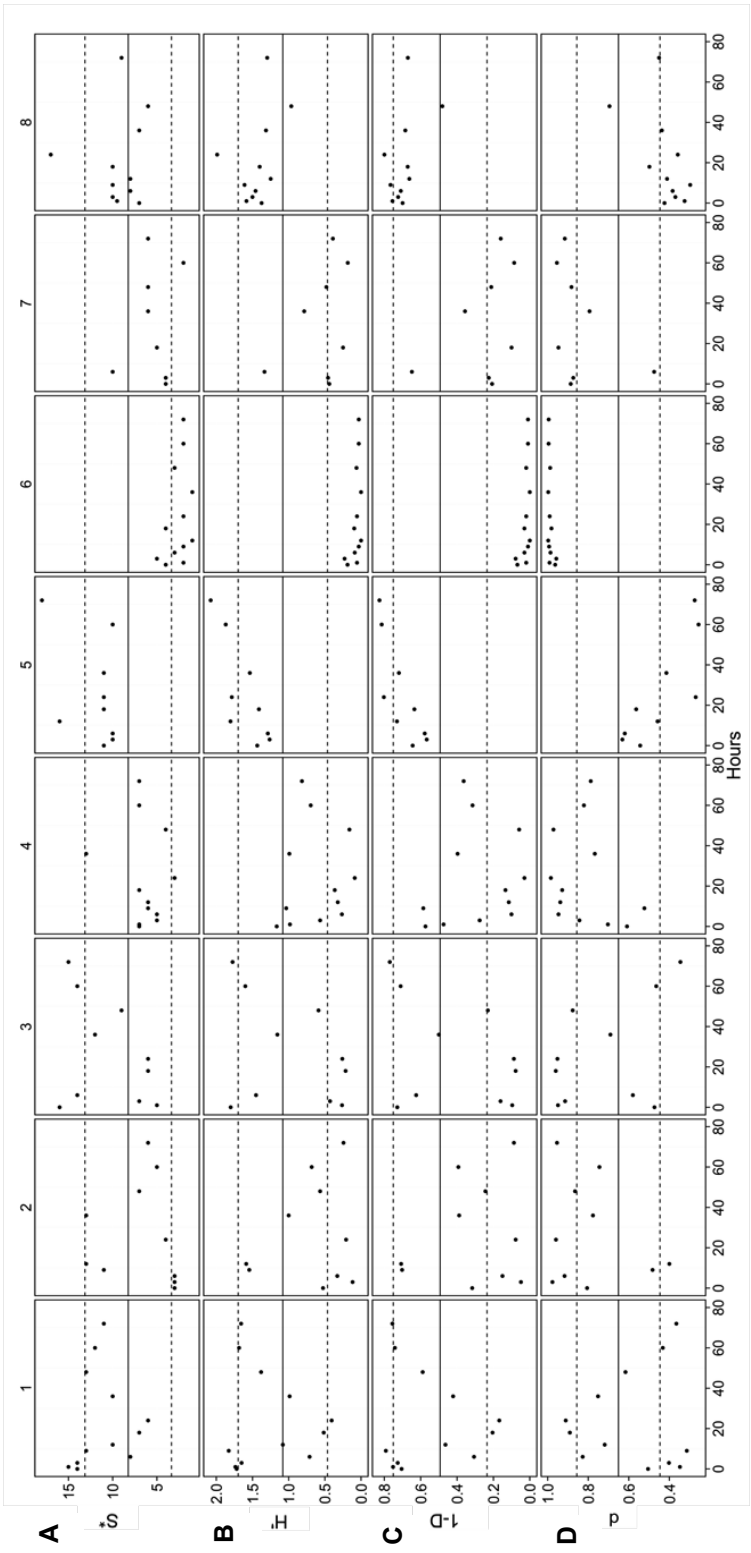


Fig. S1. Changes in diversity and dominance of bacterial communities within individual patients over time. Values of (A) species richness (*S**), (B) Shannon-Wiener index of diversity (*H'*), and (C) Simpson's index of diversity (1-*D*) are shown. The three diversity indices were calculated with a uniform re-sample size following 1000 iterations in each instance. Error bars represent the standard deviation of the mean (n = 1000). Also given is (D) the Berger-Parker index of dominance (*d*). In each instance, the overall mean (solid line) and the standard deviation of the mean (dashed lines) are shown.

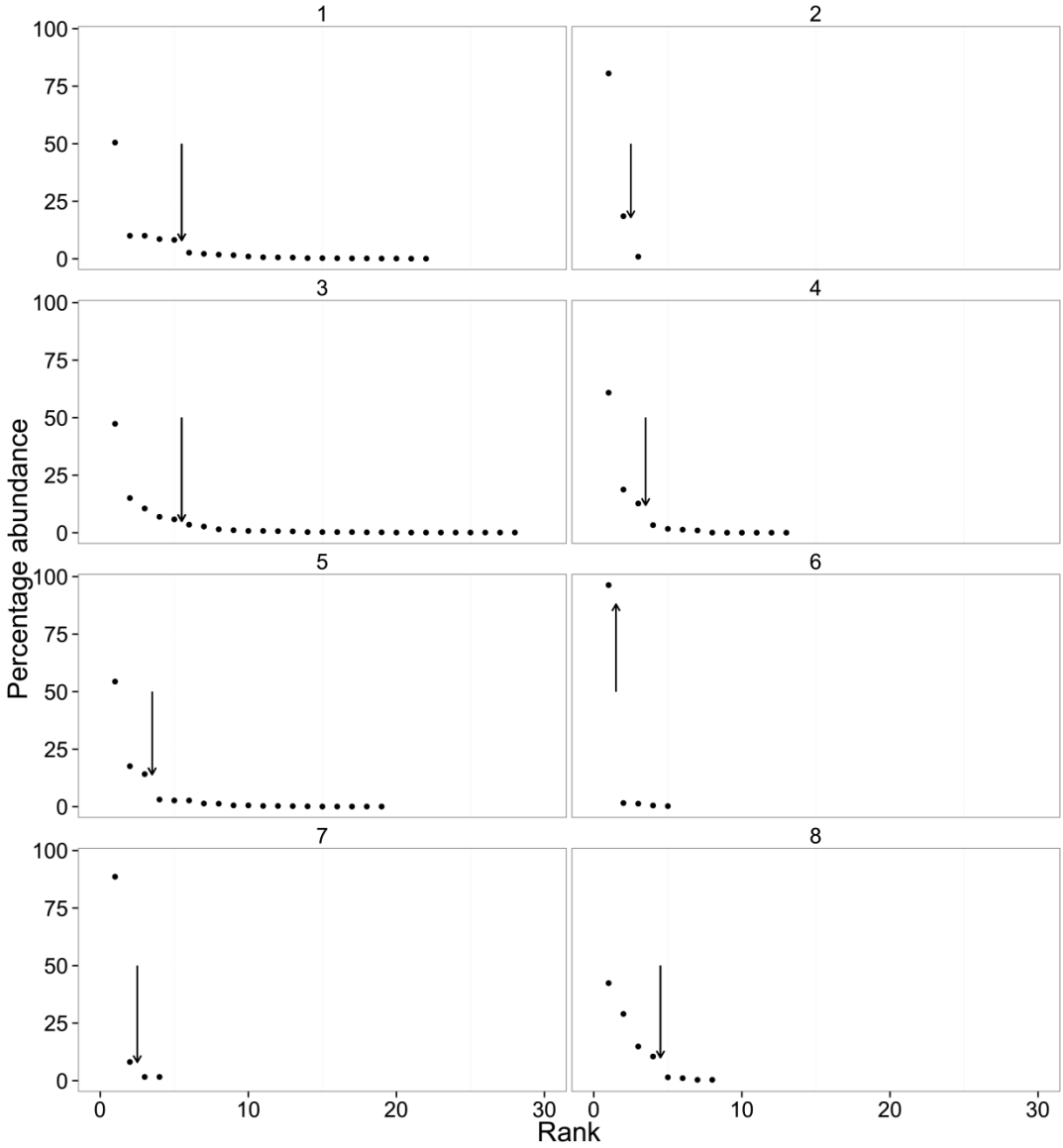


Fig. S2. Rank abundance curves for each patient bacterial community at t=0. Abundances are expressed as percentage of total abundance within each community. The arrows indicate the inflection point of each curve. Species to the left of the arrow were classified as common, and those at the right side were classified as rare.

Table S1 Bacterial species identified from eight sputum samples collected from CF patients. Species-level identities of detected taxa are reported here. However, given the length of the ribosomal sequences analysed, these identities should be considered putative. Ae denotes aerobe and An, Anaerobe. Only strict anaerobes were classified as anaerobes, whereas aerobes, facultative anaerobes, and microaerophiles were classified as aerobes. The relative abundance of the bacterial species at t=0 and t=18 are provided.

				Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Patient 7		Patient 8	
Class	**Family**	**Taxon name**	**Code**	**t=0**	**t=18**	**t=0**	**t=24**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**
Actinobacteria	Actinomycetaceae	*Actinomyces odontolyticus*	Ae	0.6	0.13	0	0	0.3	0	3.33	0.69	2.7	3.18	0	0	0	0	0	0.33
		Actinomyces oris	Ae	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0
		Actinobaculum schaalii	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Corynebacteriaceae	*Corynebacterium durum*	Ae	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Corynebacterium pseudogenitalium	Ae	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0
	Micrococcaceae	*Rothia mucilaginosa*	Ae	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33
	Propionibacteriaceae	*Propionibacterium propionicum*	Ae	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
	Bifidobacteriaceae	*Scardovia wiggsiae*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Coriobacteriaceae	*Atopobium parvulum*	An	0.3	0.06	0	0	0	0.01	1.01	0.11	1.39	1.48	0	0	0	0	0	0

Table S1 Continued

				Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Patient 7		Patient 8	
Class	**Family**	**Taxon name**	**Code**	**t=0**	**t=18**	**t=0**	**t=24**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**
Bacteroidetes	Bacteroidaceae	*Bacteroides acidofaciens*	An	0.66	0.07	0	0	0.3	0	0	0	0	0	0	0	0	0	0	0
	Porphyromonadaceae	*Barnesiella intestinihominis*	An	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0
		Porphyromonas catoniae	An	10	3.19	0.93	0	0.2	0	1.35	0.46	0.08	0	0	0.03	0	0	0	0
		Porphyromonas endodontalis	An	0	0	0	0	0	0.04	0.03	0	0.16	0	0	0	0	0	0	0
		Tannerella forsythia	An	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Prevotellaceae	*Prevotella denticola*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Prevotella enoeca	An	0	0.02	0	0	6.93	0.53	0	0	0	0	0	0.03	0	0	0	0
		Prevotella loescheii	An	0	0	0	0	0	0	0	0.11	0	0	0	0.06	0	0	0	0
		Prevotella melaninogenica	An	50.5	3.15	0	0	3.52	0.2	60.9	1.89	54.4	56.4	0.26	0.25	1.63	0.28	1.45	0.99
		Prevotella nigrescens	An	1.57	0.24	0	0	2.71	0.12	0.03	0	0	0	0	0.06	0	0	0	0
		Prevotella oralis	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Prevotella oris	An	0.18	0.04	0	0	1.51	0.09	0.07	0	0	0	0	0	0	0	0	0
		Prevotella pallens	An	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0
		Prevotella pleuritidis	An	0	0	0	0	0	0	0	0	0	0.21	0	0	0	0	0	0
		Prevotella salivae	An	2.18	0.17	0	0	0.1	0.04	0.03	0	0.57	0	0	0.03	0	0	14.9	14.5
		Prevotella veroralis	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Prevotella maculosa	An	0.18	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Prevotella tannerae	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Rikenellaceae	*Alistipes finegoldii*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S1 Continued

				Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Patient 7		Patient 8	
Class	**Family**	**Taxon name**	**Code**	**t=0**	**t=18**	**t=0**	**t=24**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**
Flavobacteria	Flavobacteriaceae	*Capnocytophaga granulosa*	Ae	1.03	0.04	0	0	0	0	0	0.06	0	0	0	0	0	0	0	0
		Capnocytophaga ochracea	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Capnocytophaga sputigena	Ae	1.81	0.13	0	0	0	0	0	0.06	0	0	0	0	0	0.28	0	0
Bacilli	Bacillales Family XI	*Gemella morbillorum*	Ae	0	0	0	0	0.6	0.04	0	0.06	0	0	0	0	0	0	0	0
	Staphylococcaceae	*Staphylococcus aureus*	Ae,P	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0.99
	Aerococcaceae	*Abiotrophia defectiva*	Ae	0	0	0	0	0.1	0	0	0	0.25	0.42	0	0	0	0	0	0
	Carnobacteriaceae	*Granulicatella adiacens*	Ae	0.24	0.02	0	0	0.1	0.01	0	0	0.57	0.42	0	0	0	0	0	0
	Lactobacillaceae	*Lactobacillus acidophilus*	Ae	0	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0
		Lactobacillus fermentum	Ae	0	0	0	0	0	0	0	0	2.7	10.4	0	0	0	0	0	0.33
		Lactobacillus johnsonii	Ae	0	0	0	0	0	0	0	0	0.08	0.42	0	0	0	0	0	0
		Lactobacillus oris	Ae	0	0	0	0	0	0	0	0	0.08	0.21	0	0	0	0	0	0
		Lactobacillus salivarius	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.09	0.33
		Lactobacillus casei	Ae	0	0	0	0	0	0	0.03	0	0.33	0.42	0	0	0	0	0	0
	Streptococcaceae	*Streptococcus mitis group*	Ae,P	10	2.32	0	0.45	10.6	1.13	12.7	1.89	14.2	16.7	1.59	0.37	1.63	0.28	42.4	49.8
Clostridia	Clostridiaceae	*Clostridium cocleatum*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Clostridiales Family XI	*Finegoldia magna*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Clostridiales Family XI	*Parvimonas micra*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Clostridiales Family XIII	*Mogibacterium diversum*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S1 Continued

				Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Patient 7		Patient 8	
Class	**Family**	**Taxon name**	**Code**	**t=0**	**t=18**	**t=0**	**t=24**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**
	Eubacteriaceae	*Eubacterium brachy*	An	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
	Lachnospiraceae	*Oribacterium sinus*	An	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0
		Stomatobaculum longum	An	0	0	0	0	0.2	0.01	0	0	0	0	0	0	0	0	0	0
		Lachnoanaerobaculum orale	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Peptococcaceae	*Peptococcus niger*	An	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Negativicutes	Veillonellaceae	*Dialister pneumosintes*	An	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
		Megasphaera micronuciformis	An	0.12	0	0	0	0.3	0.04	0	0.11	0	0.21	0	0	0	0.28	0	0.33
		Selenomonas noxia	An	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0
		Veillonella parvula	An	2.66	0.44	0	0	0.8	0.02	0	0.17	17.6	1.69	0.53	0.12	0	0.28	0.36	1.32
Fusobacteria	Fusobacteriaceae	*Fusobacterium necrophorum*	An	0	0	0	0	0.3	0	0	0	0	0	0	0	0	0	0	0
		Fusobacterium nucleatum	An	0.12	0	0	0.91	15.1	1.09	0.03	0.11	0.33	0	0	0.47	0	0	0	0.33
		Leptotrichia buccalis	An	0.3	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Leptotrichiaceae	*Sneathis sanguinegens*	An	0	0	0	0	0.7	0.07	0	0	0	0	0	0	0	0.28	0.36	0
Betaproteobacteria	Alcaligenaceae	*Achromobacter xylosoxidans*	Ae,P	8.58	0.81	80.6	95.9	5.83	0.36	1.7	1.43	1.31	7.42	1.32	0.28	88.6	94.7	29	22.8
	Burkholderiaceae	*Burkholderia cepacia complex member*	Ae,P	0	0	0	0	1.11	0	0	0	0	0	0	0	0	0	0	0
		Ralstonia pickettii	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Comamonadaceae	*Comamonas testeroni*	Ae	0	0	0	0.45	0	0	0	0	0	0	0	0	0	0	0	0
	Burkholderiaceae	*Lautropia mirabilis*	Ae	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S1 Continued

				Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Patient 7		Patient 8	
Class	**Family**	**Taxon name**	**Code**	**t=0**	**t=18**	**t=0**	**t=24**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**	**t=0**	**t=18**
	Neisseriaceae	*Eikenella corrodens*	Ae	0.06	0	0	0	0.8	0.1	0	0	0	0	0	0.06	0	0	0	0
		Kingella oralis	Ae	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
		Neisseria bacilliformis	Ae	0.54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gammaproteobacteria	Enterobacteriaceae	*Escherichia coli*	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Pasteurellaceae	*Haemophilus influenzae*	Ae,P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Haemophilus parainfluenzae	Ae	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
	Moraxellaceae	*Acinetobacter johnsonii*	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Pseudomonadaceae	*Pseudomonas aeruginosa*	Ae,P	8.22	89	18.5	2.27	47.3	96	18.8	92.8	3.11	0.42	96.3	98.1	8.13	3.64	10.5	7.26
		Pseudomonas fluorescens	Ae	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0
		Pseudomonas putida	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33
		Pseudomonas mendocina	Ae	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0
	Xanthomonadaceae	*Stenotrophomonas maltophillia*	Ae,P	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0	0	0
Mollicutes	Mycoplasmataceae	*Mycoplasma saliverium*	Ae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S2. The raw sequence data reported in this paper have been deposited in the NCBI Short Read Archive database (Accession number. SRP036061). A list of barcodes used and their associated sample are listed below.

Barcode	Patient	Hour
AGATGTTCTGCT	**1**	0
AGCTATCCACGA	**1**	1
AGTGCGATGCGT	**1**	3
ATCTGAGCTGGT	**1**	6
AGTTAGTGCGTC	**1**	9
ATACAGAGCTCC	**1**	12
AGAACACGTCTC	**1**	18
ATCCGATCACAG	**1**	24
ATCTACTACACG	**1**	36
AGCTCCATACAG	**1**	48
ATCACGTAGCGG	**1**	60
ATCTGGTGCTAT	**1**	72
ATATGCCAGTGC	**2**	0
ATCCTCAGTAGT	**2**	3
ATACACGTGGCG	**2**	6
ATCTCTGGCATA	**2**	9
ATCGTACAACTC	**2**	12
ATCAGGCGTGTG	**2**	24
ATGGATACGCTC	**2**	36
ATGGCAGCTCTA	**2**	48
ATGCACTGGCGA	**2**	60
ATGCAGCTCAGT	**2**	72
ACCACATACATC	**3**	0
ACATGTCACGTG	**3**	1
ACCTGTCTCTCT	**3**	3
ACTGATCCTAGT	**3**	6
ACGAGTGCTATC	**3**	18
ACGCTCATGGAT	**3**	24
AGACCGTCAGAC	**3**	36
ACAGTTGCGCGA	**3**	48
ACACTGTTCATG	**3**	60
ACAGCTAGCTTG	**3**	72
ACTACGTGTGGT	**4**	0
AGTCTCGCATAT	**4**	1
AGACGTGCACTG	**4**	3
ATAATCTCGTCG	**4**	6
ACGCAACTGCTA	**4**	9
AGCTTGACAGCT	**4**	12
AGGTGTGATCGC	**4**	18
AGAGAGCAAGTG	**4**	24
AGCGCTGATGTG	**4**	36
AGAGCAAGAGCA	**4**	48
ACGTGAGAGAAT	**4**	60
AGTGTTCGATCG	**4**	72
AGTACGCTCGAG	**6**	0
ATGAGACTCCAC	**6**	3
AGTGTCACGGTG	**6**	6
AGTGAGAGAAGC	**6**	12
ATATCGCTACTG	**6**	18
ATCGCGGACGAT	**6**	24
ATAGGCGATCTC	**6**	36
ATGCCTGAGCAG	**6**	60
ATGACTCATTCG	**6**	72
ACGGATCGTCAG	**7**	0
ACGCTATCTGGA	**7**	1
AGTACTGCAGGC	**7**	3
AGATCGGCTCGA	**7**	6
AGCACACCTACA	**7**	9
ACGTCTGTAGCA	**7**	12
AGCGAGCTATCT	**7**	18
ACTCACGGTATG	**7**	24
AGCAGCACTTGT	**7**	36
AGTCTACTCTGA	**7**	48
ACTGTGACTTCA	**7**	60
ACTCAGATACTC	**7**	72
ACGTACTCAGTG	**9**	0
AGCACGAGCCTA	**9**	3
AGCCATACTGAC	**9**	6
ACTAGCTCCATA	**9**	18
ACGCGCAGATAC	**9**	36
ACTTGTAGCAGC	**9**	48
ATCGATCTGTGG	**9**	60
ACGTTAGCACAC	**9**	72
ACACTAGATCCG	**11**	0
ACAGAGTCGGCT	**11**	1
ACGCGATACTGG	**11**	3
ACTGACAGCCAT	**11**	6
ACCGCAGAGTCA	**11**	9
ACAGCAGTGGTC	**11**	12
ACCAGACGATGC	**11**	18
ACCTCGATCAGA	**11**	24
ACGGTGAGTGTC	**11**	36
ACCAGCGACTAG	**11**	48
ACAGACCACTCA	**11**	72

Chapter 5: Effects of sputum freeze-thawing on CF mictobiota

Implications of multiple freeze-thawing on respiratory samples for culture independent analysis

Cuthbertson L., Rogers G.B., Walker A.W., Oliver A., Carroll M.P., Parkhill J., Bruce K.D., van der Gast C.J. (2014) **The effects of freeze thaw cycles on the microbial community present in sputum samples from the CF lung**. *Journal of Cystic Fibrosis*, doi:10.1016/j.jcf.2014.10.004

.

The use of biobanking for long term storage of respiratory samples for the use in multiple studies is increasing. It was established in the previous chapter that sample handling and storage has marked effects on bacterial community composition. As a result of these findings this chapter aimed to assess the effect of freeze-thaw cycles on the bacterial community within CF sputum. This chapter has been published in the Journal of Cystic Fibrosis.

Main title: IMPLICATIONS OF MULTIPLE FREEZE-THAWING ON RESPIRATORY SAMPLES FOR CULTURE-INDEPENDENT ANALYSES

Running title: EFFECTS OF SPUTUM FREEZE-THAWING ON CF MICROBIOTA

Leah Cuthbertson[1,2], Geraint B. Rogers[3], Alan W. Walker[4,5,], Anna Oliver[1], Lucas R. Hoffman[6,7], Mary P. Carroll[8], Julian Parkhill[4], Kenneth D. Bruce[2], Christopher J. van der Gast[1,a]

[1] *NERC Centre for Ecology & Hydrology, Wallingford, OX10 8BB, UK*

[2] *Institute of Pharmaceutical Science, Molecular Microbiology Research Laboratory, King's College London, London, SE1 9NH, UK*

[3] *SAHMRI Infection and Immunity Theme, Flinders University, Adelaide, Australia*

[4] *Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK*

[5] *Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen, AB21 9SB, UK*

[6,7] *Departments of Pediatrics and Microbiology, University of Washington, Box 356320, HSB K328A, Seattle, WA 98105, USA*

[8] *Cystic Fibrosis Unit, Southampton University Hospitals NHS Trust, Southampton, SO16 6YD, UK*

[a] To whom correspondence should be addressed: E-mail: cjvdg@ceh.ac.uk

The authors declare no conflict of interest.

Data deposition: The sequence data reported in this paper have been deposited in the NCBI Short Read Archive database (Accession number SRP040968)

Key words: Microbiota, Microbiome, Biobank, Pyrosequencing, Propidium Monoazide, Sputum

Abbreviations:

PMA: Propidium monoazide

OTU: Operational taxonomic units

*S**: Species richness

H': Shannon-Wiener

1-*D* : Simpson's

S_{BC}: Bray-Curtis

5.1 Abstract

Background

Best practice when performing culture-independent microbiological analysis of sputum samples involves their rapid freezing and storage at -80°C. However, accessing biobanked collections can mean material has been passed through repeated freeze-thaw cycles. The aim of this study was to determine the impact of these cycles on microbial community profiles.

Methods

Sputum was collected from eight adults with cystic fibrosis, and each sample was subjected to six freeze-thaw cycles. Following each cycle, an aliquot was removed and treated with propidium monoazide (PMA) prior to DNA extraction and 16S rRNA gene pyrosequencing.

Results

The impact of freeze-thaw cycles was greatest on rare members of the microbiota, with variation beyond that detected with within-sample repeat analysis observed after three cycles.

Conclusion

Four or more freeze thaw cycles results in a significant distortion of microbiota profiles from CF sputum.

5.2 Introduction

The application of next-generation sequencing technologies for the investigation of lower respiratory tract infections in patients with cystic fibrosis has revealed complex and highly diverse microbial communities [1, 2]. As technologies have improved and the associated costs have fallen, it is becoming possible to use these platforms, not only for research but also for diagnostic microbiology [3], making it more important than ever to identify and minimize the introduction of bias.

Spontaneously expectorated sputum is one of the most common specimen types used to investigate the microbial community responsible for lower respiratory infections in adults with CF. In order to perform culture-independent analysis on a representative airway sample, the methods used to collect and store specimens are hugely important. Current best practice involves the rapid stabilisation of sputum samples by freezing at -80°C within 12 hours of collection [4]. To allow profiling of the viable microbial community, samples can be treated with propidium monoazide (PMA) to remove the impact of extracellular DNA or DNA from dead and damaged cells, prior to DNA extraction and DNA sequencing [5]. Received wisdom suggests that once defrosted, respiratory samples may not be refrozen and sub-sampled again at a later date without incurring significant changes in the microbial community.

There is an increasing awareness of the importance of in-depth analysis for the investigation of the microbial communities responsible for infection. This has lead to many clinics collecting large detailed sample biobanks in ultralow temperature freezers, which can be accessed in order to address a wide range of clinical questions. However, biobanked samples may be accessed multiple times for culture independent analysis, thus passing though several freeze thaw cycles, a process that could result in changes to the microbial community. To date, no studies have used next-generation sequencing technologies to define how multiple freeze-thaw cycles affect the microbial community within collected sputum. We hypothesized that microbial community profiles would be significantly altered

with each additional freeze-thaw cycle when analysed using 16S rRNA gene pyrosequencing.

5.3 Methods

5.3.1 Sample collection

Sputum samples were collected, under full ethical approval from the Southampton and South West Hampshire Research Ethics Committee (06/Q1704/26), from eight patients attending the regional Cystic Fibrosis Centre in Southampton General Hospital. All patients were chronically colonised with *Pseudomonas aeruginosa*. Patients were selected based on their ability to typically produce more than 2 mls of sputum. Sputum samples were collected and frozen at -80°C within 1 hour. Each sputum sample was subjected to six freeze-thaw cycles. Samples were removed from the -80°C freezer, a 250 µl aliquot removed for DNA extraction, and the remaining sample allowed to completely thaw at room temperature for 30 min before being returned to -80°C for 24 hours.

5.3.2 DNA extraction and Pyrosequencing

Sputum samples were washed three times with 1x phosphate buffered saline to remove saliva, as previously described [1]. Extracellular DNA and DNA from non-viable cells were excluded from analysis via crosslinking with PMA [6, 7] prior to DNA extraction, as described previously [8]. Bacterial Golay barcode-encoded FLX amplicon pyrosequencing was performed using the primer 338F (3'- ACTCCTACGGGAGGCAGCAG) and 926R (3'- CCGTCAATTCMTTTRAGT). Initial generation of 16S rRNA gene amplicons involved a one step PCR of 25 cycles using AccuPrime^{TM} Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA). 454 pyrosequencing using the Lib-L kit was performed at the Wellcome Trust Sanger Institute, Hinxton, UK.

Resulting data were analysed using the Mothur sequencing analysis platform [9] as described previously [4]. The raw sequence data generated within the current study have been submitted to the NCBI Short Read Archive database under the study accession number SRP040968. The barcodes associated with each sample are shown in Table S1. Two aliquots were excluded due to insufficient number of sequence reads generated.

5.3.3 Statistical analysis

Statistical analysis was performed in R [10]. Changes in bacterial diversity were assessed using three complementary measures: species richness (*S**,the total number of species), Shannon-Wiener (*H*', a metric accounting for both number and relative abundance of species), and Simpson's (1-*D,* a measure of the probability that two species randomly selected from a sample will differ) indices of diversity as described previously [4, 11]. The Bray-Curtis (S_{BC}, which accounts for the number and abundance of species present in each community and those that are shared), resulting in a value between 0 and 1 (higher values indicating greater similarity) measure of similarity was used to assess changes in community composition with each freeze-thaw cycle.

To avoid potential bias, all measures were calculated using randomised resampling to a uniform number of sequence reads per sample [5]. Mean diversity measures were calculated from the re-sampling of the reads from each specimen to the lowest number of sequence reads among all specimens (n=261) for 1000 iterations. S_{BC} was calculated by re-sampling to the minimum number of sequence reads per specimen within each patient and comparing community composition to the original sample for 1000 iterations. Bacterial species detected at the first point for each patient were partitioned into common and rare species using rank abundance curves [12]. The R package nlme [13] was used to fit mixed effect models to investigate the relationships between measures of diversity, similarity, and number of freeze thaw cycles. r^2 values were calculated using the MuMIn package [14].

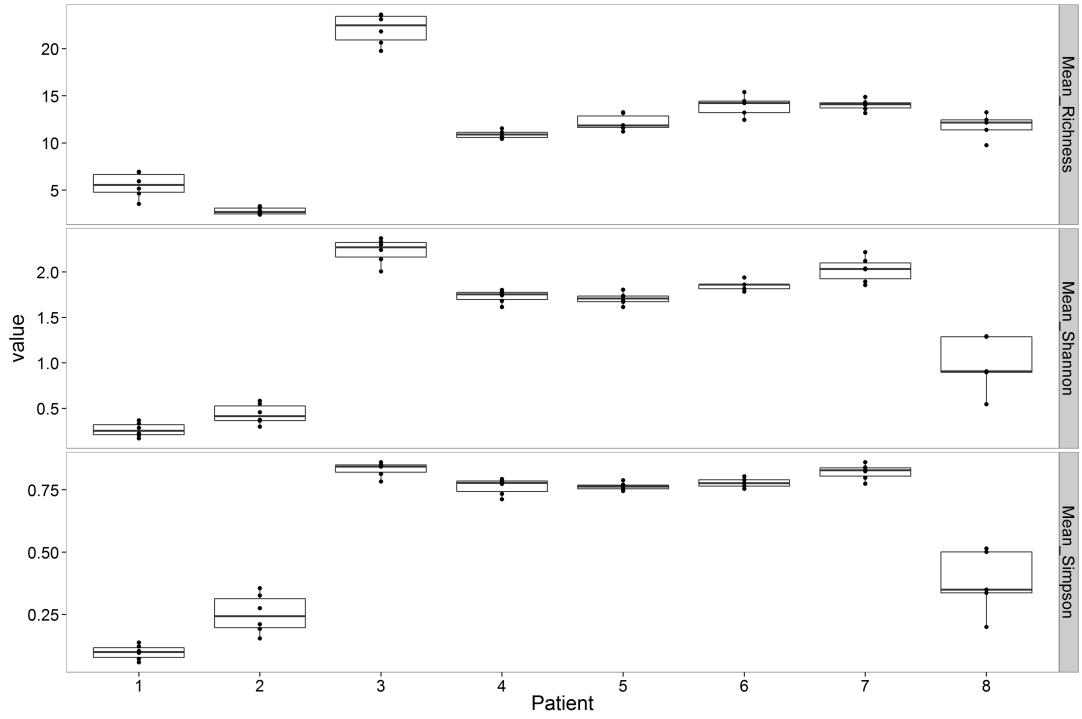
5.4 Results

To test the study hypothesis, sputum samples from eight CF patients were subjected to six freeze-thaw cycles. Aliquots of sputum were removed for DNA extraction and 16S rRNA gene pyrosequencing to assess the bacterial community after each cycle, and the remaining sample was allowed to defrost completely before being returned to -80°C. All samples were treated with PMA prior to DNA extraction to focus the analysis on the viable bacterial community. A total of 106,065 sequences (mean ± standard error (SE) per sample 2306 ±239) were generated from 46 samples, identifying 49 genera and 76 distinct operational taxonomic units (OTUs) classified to species level (Table S2).

5.4.1 Bacterial diversity

Species richness, *S**, was found to be highly variable between patients (max= 49, min=4), no pattern in raw richness values was observed over the freeze-thaw cycles either within patients (P=0.6, r^2=0.0004) or when using randomised resampling to calculate mean *S** (P =0.7, r^2= 0.0001). High levels of variability in bacterial diversity were also observed between patients as measured by *H*' and 1-*D*. However, as with *S**, no pattern in either of these diversity measures was observed over the 6 freeze-thaw cycles (*H*' (P=0.4, r^2=0.001), 1-*D* (P=0.1, r^2=0.003) (Fig 1).

Fig 1. Boxplots from each patient showing variation in diversity over the 6 freeze-thaw cycles. Values of (A) species richness (*S**), (B) Shannon-Wiener index of diversity (*H*'), and (C) Simpson's index of diversity (1-*D*) were calculated with a uniform re-sample size following 1000 iterations in each instance. Lines in boxplot represent mean and standard deviation of the mean (*n*=6); whiskers represent 25th and 75th percentiles.



5.4.2 Bacterial community membership

Changes in community composition were compared using the S_{BC} measure of similarity. A significant decrease in similarity was observed over the 6 cycles ($P < 0.0001$, $r^2 = 0.40$) (Fig 2). Variation in intra-sample composition could contribute to the effect associated with freeze-thaw cycles [4, 15]. Therefore, to account for within-sample variation, a cut off value for similarity was calculated by comparing the within-sample similarities of triplicate sequence datasets of eight samples that were published previously [4]. Based on this data, Bray-Curtis similarity values below 0.682 were judged to differ significantly from the original sample.

To minimise bias, changes in similarity were measured using randomised resampling to a uniform subsample size. Analysis showed that, despite a significant trend of decreasing similarity between aliquots within each sample over the six freeze-thaw cycles, none of these changes in within-sample similarity fell below the cut-off value of 0.682 for significant within-sample variation.

To investigate the drivers of this trend in community composition, we partitioned the community into common and rare species using rank abundance curves, indicating the breakpoint between the common and rare using the inflection point (Fig. S1) [12]. Once partitioned, the Bray-Curtis similarity was calculated for the common and rare species. The common and abundant species were most important in determining the overall trend in community similarity, while the rare species were shown to be more variable (Fig 2). Over the six cycles, the common species were not found to fall below the cut-off value for within-sample variation. In contrast, however, substantial variation was observed in the detection of rare species, for which the mean change in similarity among aliquots from each sample dropped below the expected level for within-sample variation after four freeze-thaw cycles.

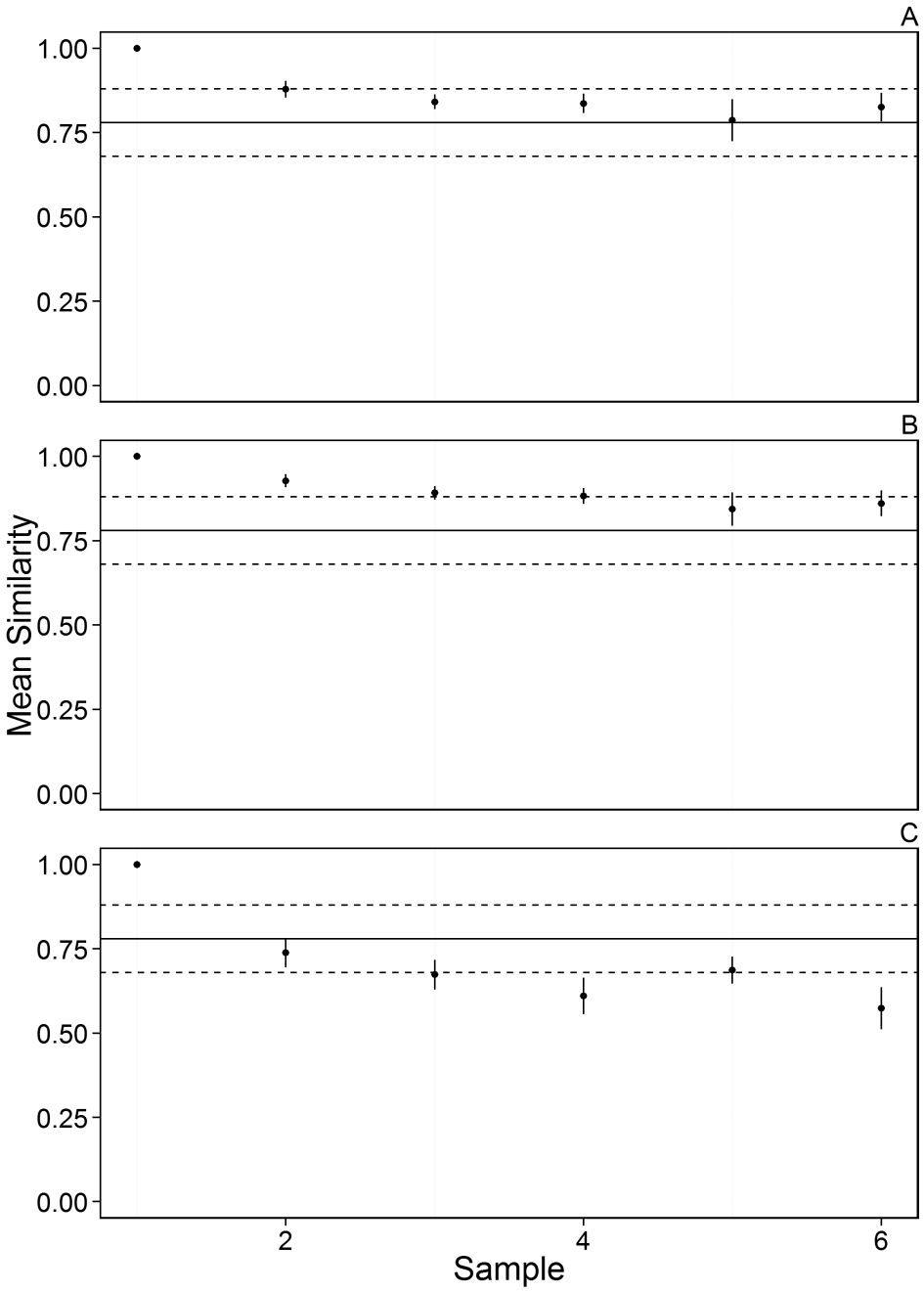


Fig 2. Mean changes in bacterial community composition over six freeze-thaw cycles, using Bray-Curtis index of similarity for (A) whole communities, (B) common and (C) rare species groups. Error bars represent the standard error of the mean (*n* = 6). Whole community similarity was calculated with a uniform re-sample size following 1000 iterations. Solid lines represent the overall mean similarity for within sample replicates and dashed lines represent the standard error of the mean (*n* = 24).

5.5 Discussion

The handling of respiratory samples for culture-independent analysis is vital if an unbiased picture of the microbial community is to be obtained. Storage of samples at -80°C is the standard recognised method for maintenance of sample integrity during biobanking. However, whether repeated sub-sampling of these specimens, which will require repeated thawing and freezing, leads to sample degradation and changes in community composition detected has not been reported.

In searching for changes in bacterial diversity due to freeze-thaw cycles, no significant overall trend was observed. A significant negative trend in community similarity was observed using the Bray-Curtis measure of similarity. Despite this observation, over the 6 cycles the change in similarity never fell below the expected level for within-sample variation, indicating that repeat freeze-thaw cycles will not affect the overall bacterial community composition. However, the rare community fell below the level of within sample variation from 4 freeze-thaw cycles.

Our results challenge the long-held view that the microbial community within respiratory samples will significantly change once samples have been defrosted during subsampling. More than three freeze-thaw cycles will result in significant divergence of the rare community from the original sample. If the sample is subjected to four or more cycles community analysis may be carried out on the common community but the rare should be interpreted with care. In practical terms, these findings support the aliquoting of samples to avoid unnecessary freeze-thaw cycles.

5.1 Acknowledgements

This study was supported by the UK Natural Environment Research Council (NE/H019456/1) to CJvdG, by the Wellcome Trust (WT 098051) to AWW and JP, and by the US National Institute of Health (K02HL105543 and P30 DK089507) to LRH.

5.2 References

[1] Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, *et al*. **Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis**. J Clin Microbiology, 44(2006), pp.2601-2604.

[2] Loens K, Van Heirstraeten L, Malhotra-Kumar S, Goossens H, Ieven M. **Optimal Sampling Sites and Methods for Detection of Pathogens Possibly Causing Community-Acquired Lower Respiratory Tract Infections**. J Clin Microbiol. 47(2009), pp.21-31.

[3] Pattison SH, Rogers GB, Crockard M, Elborn JS, Tunney MM. **Molecular detection of CF lung pathogens: Current status and future potential.** J Cyst Fibros. 12(2013), pp.194-205.

[4] Cuthbertson L, Rogers GB, Walker AW, Oliver A, Hafiz T, Hoffman LR, *et al*. **Time between sputum sample collection and storage significantly influences bacterial sequence composition from Cystic Fibrosis respiratory infections**. J Clin Microbiol. 52(2014), pp.3011-3016

[5] Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, *et al*. **Reducing bias in bacterial community analysis of lower respiratory infections**. ISME J. 7(2013), pp.697-706.

[6] Nocker A, Sossa-Fernandez P, Burr MD, Camper AK. **Use of propidium monoazide for live/dead distinction in microbial ecology**. Appl Environ Microbiol. 73(2007), pp.5111-5117.

[7] Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD. **Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections**. Diagn Microbiol Infect Dis. 62(2008), pp.133-141.

[8] Rogers GB, Marsh P, Stressmann AF, Allen CE, Daniels TV, Carroll MP, *et al*. **The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples.** Clin Microbiol Infect. 16 (2010), pp.1656-1658.

[9] Schloss PD, Gevers D, Westcott SL. **Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies**. PLOS One. 6(2011), e27310

[10] The R Core Development Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2013.

[11] van der Gast C, Walker A, Stressmann F, Rogers G, Scott P, Daniels T, *et al*. **Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities**. ISME J. 5(2011), pp.780-791.

[12] Siqueira T, Bini LM, Roque FO, Couceiro SRM, Trivinho-Strixino S, Cottenie K. **Common and rare species respond to similar niche processes in macroinvertebrate metacommunities**. Ecography. 35(2012), pp.183-92.

[13] Pinheiro J BD, DebRoy S, Sarkar D, and the R Development Core Team. **nlme: Linear and Nonlinear Mixed Effects Models**. R package version 31-110. (2013)

[14] Barton K**. MuMIn: Multi-model inference**. R package version 195. (2013)

[15] Zhao JC, Li J, Schloss PD, Kalikin LM, Raymond TA, Petrosino JF, *et al*. **Effect of Sample Storage Conditions on Culture-Independent Bacterial Community Measures in Cystic Fibrosis Sputum Specimens**. J Clin Microbiol. 49(2011), pp.3717-3718.

5.3 Supplementary information

Fig S1. Rank abundance curves for each patient's bacterial community after a single freeze thaw cycle. Abundances are expressed as percentage of total abundance within each community. The arrows indicate the inflection point of each curve. Species to the left of the arrow were classified as common, and those at the right side were classified as rare.

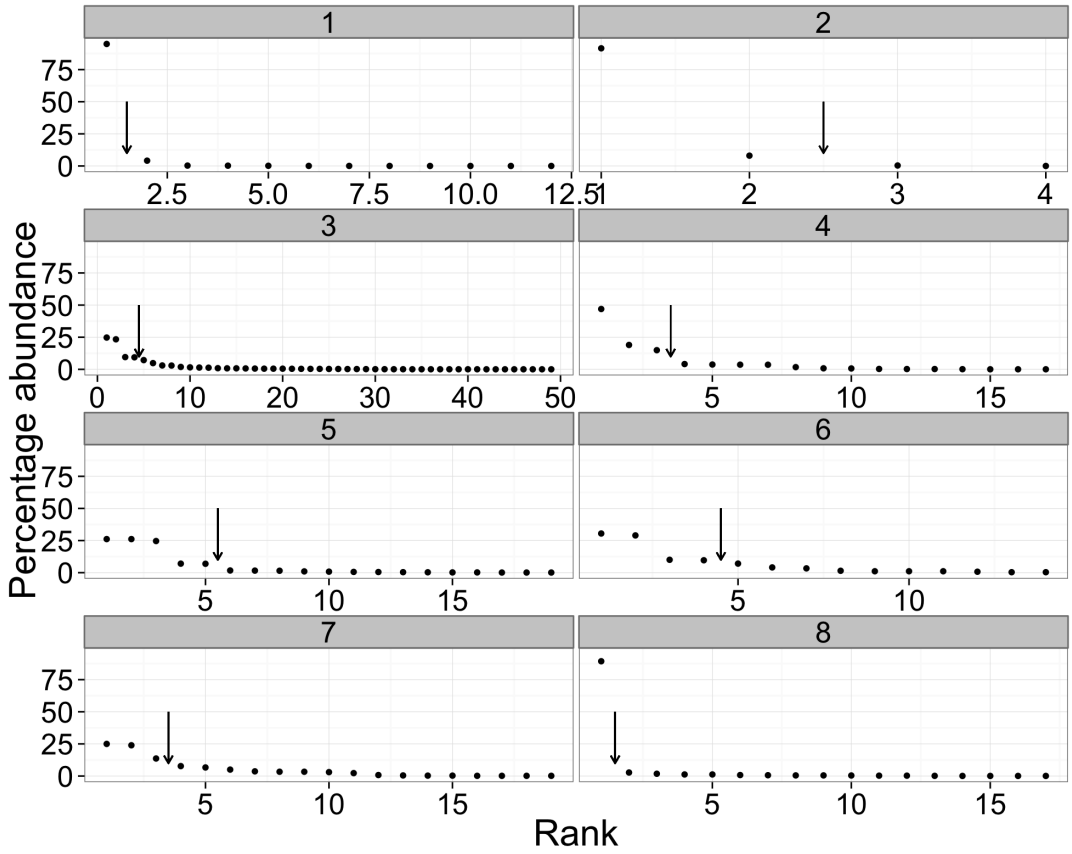


Table S1. The raw sequence data reported in this paper have been deposited in the NCBI Short Read Archive database (Accession number PRJNA243349). A list of barcodes used and their associated sample are listed below.

Barcode	Patient	Sub-sample
ATGGCGTGCACA	1	1
CAACACGCACGA	1	2
GTCTATCGGAGT	1	3
GTGAGGTCGCTA	1	4
GTGTGTGTCAGG	1	5
TAAGCGCAGCAC	1	6
ATGGTCTACTAC	2	1
CAACTATCAGCT	2	2
GTCTCATGTAGG	2	3
GTGATAGTGCCG	2	4
GTGTTGCAGCAT	2	5
TACACACATGGC	2	6
ATGTACGGCGAC	3	1
CAACTCATCGTA	3	2
GTCTCTCTACGC	3	3
GTGCAATCGACG	3	4
GTTAGAGCACTC	3	5
TACACGATCTAC	3	6
ATGTCACCGTGA	4	1
CAAGATCGACTC	4	2
GTCTGACAGTTG	4	3
GTGCACATTATC	4	4
GTTCGCGTATAG	4	5
TACAGATGGCTC	4	6

Barcode	Patient	Sub-sample
ATGTGCACGACT	5	1
GTCGCTGTCTTC	5	2
GTCTGGATAGCG	5	3
GTGGCGATACAC	5	4
GTTGACGACAGC	5	5
TACAGTCTCATG	5	6
ATGTGTCGACTT	6	1
GTCGTAGCCAGA	6	2
GTCTTCGTCGCT	6	3
GTGTACCTATCA	6	4
GTTGTATACTCG	6	5
ATTATCGTGCAC	7	1
GTCGTGTGTCAA	7	2
GTGACCTGATGT	7	3
GTGTCTACATTG	7	4
TAACAGTCGCTG	7	5
ATTCTGTGAGCG	8	1
GTCTACACACAT	8	2
GTGACTGCGGAT	8	3
GTGTGCTATCAG	8	4
TAACTCTGATGC	8	5
TACGATGACCAC	8	6

Table S2. Bacterial taxa samples across eight sputum samples from adult CF patients. Species-level identities of detected taxa are reported here. However, given the sequence read length of the ribosomal sequences analysed, these identities should be considered putative

Class	Family	Taxon name
Actinomycetales	Actinomycetaceae	*Actinomyces odontolyticus*
		Actinomyces oris
	Corynebacteriaceae	*Corynebacterium durum*
		Corynebacterium matruchotii
	Micrococcaceae	*Rothia mucilaginosa*
	Propionibacteriaceae	*Propionibacterium propionicum*
Bacillales	Bacillales	*Gemella morbillorum*
	Staphylococcaceae	*Staphylococcus aureus*
Bacteroidales	Bacteroidaceae	*Bacteroides acidofaciens*
		Bacteroides vulgatus
	Porphyromonadaceae	*Porphyromonas catoniae*
		Porphyromonas endodontalis
		Tannerella forsythia
	Prevotellaceae	*Alloprevotella rava*
		Prevotella denticola
		Prevotella histicola
		Prevotella loescheii
		Prevotella maculosa
		Prevotella melaninogenica
		Prevotella nanceiensis
		Prevotella nigrescens
		Prevotella oralis
		Prevotella oris
		Prevotella pallens
		Prevotella pleuritidis
		Prevotella salivae
		Prevotella tannerae
	Rikenellaceae	*Alistipes finegoldii*

Table S2 Continued

Class	Family	Taxon name
Burkholderiales	Alcaligenaceae	*Achromobacter xylosoxidans*
	Burkholderiaceae	*Lautropia mirabilis*
	Sutterellaceae	*Sutterella wadsworthensis*
Campylobacterales	Campylobacteraceae	*Campylobacter concisus*
Clostridiales	Clostridiaceae	*Clostridium* sp.
	Clostridiales	*Parvimonas micra*
	Eubacteriaceae	*Eubacterium brachy*
		Eubacterium sulci
	Lachnospiraceae	*Butyrivibrio fibrisolvens*
		Catonella morbi
		Lachnoanaerobaculum orale
		Oribacterium sinus
		Shuttleworthia satelles
		Stomatobaculum longum
	Peptostreptococcaceae	*Peptostreptococcus stomatis*
Coriobacteriales	Coriobacteriaceae	*Atopobium parvulum*
		Olsenella uli
Enterobacteriales	Enterobacteriaceae	*Escherichia coli*
Flavobacteriales	Flavobacteriaceae	*Capnocytophaga granulosa*
		Capnocytophaga haemolytica
		Capnocytophaga ochracea
		Fusobacterium nucleatum
	Leptotrichiaceae	*Leptotrichia buccalis*
		Leptotrichia wadei
		Sneathia sanguinegens
Lactobacillales	Aerococcaceae	*Abiotrophia defectiva*
	Carnobacteriaceae	*Granulicatella adiacens*
	Lactobacillaceae	*Lactobacillus fermentum*
		Lactobacillus johnsonii
	Streptococcaceae	*Streptococcus mutans*
		Streptococcus parasanguinis
		Streptococcus salivarius
Mycoplasmatales	Mycoplasmataceae	*Mycoplasma salivarium*
Neisseriales	Neisseriaceae	*Kingella denitrificans*
Pasteurellales	Pasteurellaceae	*Haemophilus parainfluenzae*
Pseudomonadales	Moraxellaceae	*Acinetobacter johnsonii*
		Moraxella osloensis

Table S2 Continued

Class	Family	Taxon name
	Pseudomonadaceae	*Pseudomonas aeruginosa*
		Pseudomonas putida
Rhizobiales	Hyphomicrobiaceae	*Hyphomicrobium sp.*
Selenomonadales	Veillonellaceae	*Selenomonas noxia*
		Anaeroglobus geminatus
		Dialister invisus
		Megasphaera micronuciformis
		Selenomonas artemidis
		Selenomonas noxia
		Selenomonas sputigena
		Veillonella parvula

Chapter 6: The complex relationships between the bacterial community and clinical factors in cystic fibrosis

The complex relationships between the bacterial community and clinical factors in cystic fibrosis

6.1 Introduction

In recent years, through a range of culture independent techniques, the complex and diverse nature of the bacterial community associated with cystic fibrosis (CF) lung infections has been recognised (1-6). However despite this insight, chronic infection and concomitant inflammation are still the major cause of morbidity and mortality in individuals with CF (7). Therefore, in order to improve patient management it is important to increase our understanding of the bacterial community within the CF lung in relation to clinical factors for example; gender, age, body mass index (BMI), lung function, liver disease, diabetes, pancreatic insufficiency and genotype.

Over the last few years, the importance of relating clinical outcomes to changes in the bacterial community has been recognised. In 2010, Klepac-Ceraj *et al* examined the influence of clinical factors on the bacterial community present in 45 CF children using PhyloChip hybridization (8). This study examined the relationship between the community composition and clinical factors including age, CFTR genotype, and antibiotic therapy; as well as examining the effect of *Pseudomonas aeruginosa* colonisation (8). While this study found no relationship between bacterial diversity and lung function, bacterial diversity was found to decrease with age (8). Further, a study by Cox *et al* (2010) used an age stratified patient cohort to reveal a significant negative relationship between age and lung function of patients from 0 -72 years old. They identified that older patients tended to have a more conserved bacterial community than younger individuals (3).

The lack of association of the bacterial community with lung function, observed by Klepac-Ceraj *et al* (2010), was considered to be a result of CF children having lung function scores comparable with healthy individuals. Therefore, it was unsurprising that in a study of 14 adult CF patients, van der Gast *et al* (2011) using full length 16S rRNA gene sequencing to investigate the metacommunity within the CF lung, revealed a significant positive relationship between bacterial diversity and lung function (9). This result was also observed in a study by Delhaes *et al* (2012) who found that bacterial diversity was correlated with lung function and

poor clinical status (6). Despite the significant relationship observed, the variance associated with this relationship was shown to be high in both these studies (6, 9). As a result, to establish if this relationship could be used in order to predict clinical outcomes a much larger sample set would be required (6).

While the effect of relationship between the bacterial community and lung function is clearly important, it cannot account for all the variance observed between patients, so other clinical factors must be considered. It has been established that different cystic fibrosis transmembrane regulator (CFTR) genotypes result in different clinical outcomes for CF patients (10). By grouping patients into high and low risk CFTR mutation categories, McKone *et al* (2006) found significantly different survival rates as a result of mutation type (11). ΔF508 is the most common CFTR mutation associated with CF and is classified as one of the severe mutations associated with disease. In order to investigate if differences in CFTR genotype influence the bacterial community within the CF lung, Klepac-Ceraj *et al* (2010) partitioned their samples into 3 distinct groups, ΔF508 homozygote's, ΔF508 heterozygote's and other non-ΔF508 mutations (8). Their results suggested that the ΔF508 mutation may have a marked effect on the environment within the CF lung, resulting in a distinct bacterial composition (8). However, they do accept that a larger cohort study would be required to investigate this relationship further (8).

Several large prospective studies have demonstrated that there is a gender gap in CF, resulting in significantly poorer prognosis for female CF patients compared to males (11, 12). Early studies showed that between the ages of 1 and 20 years, females were 60% more like to die than males (12). With improvements in treatment regimes and increase in patient survival rates, Verma *et al* (2006) investigated if differences in disease prognosis were still an issue despite advances in patient care (13). This study concluded that with modern treatments, no significant difference was observed between the lung function of male and female patients however, females tended to show poorer growth rates than males. This study did not investigate changes in the bacterial community between the two sexes and

therefore was unable to uncover if the differences in growth rate observed were related to patient gender.

The course of disease progression may also be affected by the centre attended by an individual patient. This could be explained by differences in treatment regimens or simply by the environmental microbiota present in different geographical regions. In 1998, Johansen *et al* undertook a study to investigate differences in clinical status between two CF centres, one in Toronto, Canada and the other in Copenhagen, Denmark (14). This study found that despite differing treatment regimes between centres, pulmonary and nutritional statuses between the two cohorts was not significantly different (14). A similar study was carried out by Stressmann *et al* (2011), using T-RFLP to investigate differences in bacterial community composition between samples collected from Southampton, United Kingdom and those originating from Chapel Hill, United States (15). This study revealed geographical differences between the presence and relative abundance of species identified (15). The most common organisms identified were found to be common to both sites (15) however, while this study showed that the community present in the US samples were less diverse than those in the UK, changes in the metacommunity between centres was not investigated.

As indicated above, although some evidence has been obtained linking the bacterial community to clinical factors the studies have been limited by participation. Typically, these studies have been carried out using samples from a small number of patients (ca. less than 50 patients) attending a single CF centre. In this study over 300 sputum samples were collected from 11 different sites across Europe and North America, making this the largest study of its kind, and resulting in statistical power unrivalled by the currently published work. This dataset therefore results in a unique opportunity to test the relationships previously observed in much smaller datasets and whether previously unobserved associations can be revealed.

6.2 Materials and methods

6.2.1 Sample collection

Sputum samples were collected from CF patients attending CF clinics across Europe and the United States of America. For full details of clinics and associated ethics see Table 2.2. Single samples were collected from each patient alongside associated clinical metadata. Once collected, samples were stored at -80°C prior to transport on dry ice to the Centre for Ecology and Hydrology in Wallingford. Clinical metadata associated with each sample was collected at the time of sampling, a summary is shown in Table 6.1.

A total of 379 individual samples were sequenced. After processing in Mothur a total of 44 technical control samples, including reference strains (shown in Table 2.1), were used to remove contaminating OTUs. All controls used are shown in Table 6.2. Due to a lack of sequences, or as a consequence of contamination, 24 samples were removed from the final analysis. As a result, a total of 292 CF sputum samples were included in the analysis of this dataset. Additionally, 19 non-CF healthy controls, sampled by sputum induction using inhalation of hypertonic saline, at Belfast city hospital were also included.

Table 6.1 Patient data and samples.

Variable		***n***
Non-CF healthy controls	9 male	10 female
Patients		**292**
Sex	170 male	122 female
BMI, *n* [min-max]	183	[14.4-43.8]
FEV_1, *n* [min-max]	276	[0.43-4.93]
Age, *n* [min-max]	254	[14-71]
CFTR genotype, *n* (%)		**252**
ΔF508/ΔF508	135	(54)
ΔF508/other	98	(39)
other/other	19	(8)
Clinical Status, *n* (%)*		**216**
Stable	162	(75)
Exacerbation	54	(25)
Pancreatic sufficient, *n* (%)		**225**
Sufficient	86	(38)
Insufficient	139	(62)
Liver disease, *n* (%)		**225**
Yes	37	(16)
No	188	(84)
Diabetes, *n* (%)		**287**
Yes	98	(34)
No	189	(66)
Location, n (%)		**292**
Belfast, UK	60	(21)
Southampton, UK	35	(12)
London, UK	14	(5)
Dublin, Eire	6	(2)
Warsaw, Poland	17	(6)
Dartmouth, Bedford, NH	17	(6)
Dartmouth, Lebanon, NH	4	(1)
Portland, ME	22	(8)
Boston, MA	15	(5)
Seattle, WA	75	(26)
Vermont, ME	27	(9)

Bold text indicate the total number of samples with available metadata relating to particular clinical variables, square brackets indicate the sample range, round brackets indicate the % of the total number of samples with data for the specified variable. * Patients diagnosed and in current treatment for a CF pulmonary exacerbation (CFPE) were defined as having a clinical status of "Exacerbation".

Table 6.2 Technical control samples included in sequencing analysis.

Mock communites contained equal quanties of *Achromobacter xylosoxidans, Burkholderia cenocepacia, Burkholderia multivorans, Pseudomonas aeruginosa, Staphylococcus aureus* and *Stenotrophomonas maltophilia*. Extraction controls were carried out using sterile water during every extraction protocol and sequenced to identify extraction contaminants. A mock community was included in each run, the PCR controls were carried on sterile water.

Controls	***n***
Achromobacter xylosoxidans	3
Burkholderia cenocepacia	3
Burkholderia multivorans	3
Pseudomonas aeruginosa	3
Staphylococcus aureus	3
Stenotrophomonas maltophilia	3
Mock communities	9
Extraction controls	14
PCR controls	3

6.2.2 DNA extraction and pyrosequencing

All samples were stored at -80°C until processing. A sterile scalpel was used to transfer approximately 250µl of frozen sputum into a 15ml centrifuge tube, as described in Chapter 2.3.1, to allow a sputum wash to be performed. Using the method described in Chapter 2.3.2, washed sputum was then treated with propidium monoazide (PMA) prior to DNA extraction; Chapter 2.3.3.

Extracted DNA was amplified using modified primers 27F (5'-AGMGTTYGATYMTGGCTCAG) (MWG Eurofins, Ebersberg, Germany) and 338R (5'-GCTGCCTCCCGTAGGAGT) (MWG Eurofins, Ebersberg, Germany) for 16S rRNA gene sequencing using Illumina MiSeq, and as described in Chapter 2.4.2. 16S rRNA gene amplicons were initially generated using a one step PCR of 25 cycles using Q5® High-Fidelity DNA Polymerase (New England Biolabs, Ipswitch, UK). Illumina MiSeq sequencing was performed at the Wellcome Trust Sanger Institute, Hinxton, UK.

6.2.3 Sample processing and sequence analysis

Resulting data was analysed using the Mothur sequencing analysis platform using a modified version of the method described in Chapter 2.7. Before analysis the two sets of output reads for each sample were combined to create contigs. Sequences shorter than 270 bases and longer than 400 bases were removed, along with any sequences that included ambiguous base calls and homopolymeric stretches longer than 7 bases. Any non-bacterial ribosome sequences or chimeras were removed using Chimera uchime software, as implemented in Mothur (16). Sequences were classified using the Ribosomal database project (RDP) training set version 9 and any lineage identified as chloroplast, mitochondria, unknown, archaea or eukaryote were removed (17). Prior to creating a distance matrix and clustering OTUs, each barcoded sample was subsampled to 2000 sequences. Sequences were then assembled into operational taxonomic units (OTUs) (clustered at 97% identity), to give an approximation of species (18), and identified using the RDP reference database version 9 (2012) (17). Representative sequences were used to give an appropriate species

level identification for the OTU using NCBI multiblast and any suspect OTUs, those that appeared in control samples as well as single sequences identified as being common contaminants (e.g. hyperthermophiles, methanotrophs and rhizosome associated bacteria), were removed (19).

6.2.4 Statistical analysis

Statistical analysis was performed in R version 3.1.1(2012-07-10) (20), PAST (version 2.17) and XLSTAT (version 2014.3, Addinsoft, USA).

Species were partitioned using a distribution abundance relationship (DAR), described by Magurran and Henderson (2003) (21). The most persistent and abundant species, those present in more that 75% of the total samples, were described as common, while all others were considered to be rare.

Three recognised measures of bacterial diversity were used to assess bacterial richness and evenness; species richness (*S**), Shannon-Wiener diversity index (*H'*) and Simpson's index of diversity (1-*D*), as described in Chapter 2.81. Randomised resampling was carried out while assessing diversity measures in order to standardise the effect of variable read numbers returned by sequencing each sample. Diversity measures were calculated with the resulting data using the vegan package in R (22). The relationship between diversity and continuous metadata was assessed using linear regression in R. Alternatively analysis of variance (ANOVA) was used to assess differences in categorical variables and Tukey HSD tests were used to compare differences between categories.

Community composition was investigated using Sørensen and Bray-Curtis measures of similarity, see Chapter 2.8.2 for full details. Analysis of similarity (ANOSIM) was used to assess the similarity between continuous variables, using PAST (version 2.7) (23). This method uses similarity measures to investigate similarity within variables and compares the allowing differences to be assessed. *R* values are on a scale of +1 to -1; values of +1 indicating the most similar samples are within the same group, and -1 that the most similar samples are outside the group. Significant *R* values indicate that the community similarities

are more similar within the disease period and therefore can be considered significantly different.

The contribution of each OTU to the observed Bray-Curtis similarity between significant variables was assessed using Similarity of Percentages (SIMPER) analysis. This was done in order to identify those OTUs that are important in creating the observed pattern of similarity. SIMPER analyses were performed as previously described by Clarke (1993) (24), using PAST (version2.7) (23).

Mantel tests were used to relate the variability in bacterial community similarity to the clinical factors shown in Table 5.1 (25). Bray-Curtis similarity matrices were created in using PAST (version 2.7), and similarity matrices were generated by calculating absolute differences between variables. Similarity matrices for categorical data were created by assigning 0 to samples which were in the same category and 1 to those that were different, this avoided any weighting associated with the categories. Lower tailed Mantel and partial Mantel tests were conducted using XLSTAT (version 2014.3) with *P*-values based on 9999 permutations (9, 26).

6.3 Results

Sputum samples collected from CF patients across Europe and North America were collected with complementary clinical information. The bacterial diversity and community composition was assessed for each sample by 16S rRNA gene sequencing using Illumina MiSeq. A total of 311 sputum samples were included in this analysis. Each sample was subsampled to 2000 sequences in Mothur and any contaminating OTUs were removed post Mothur analysis. As a result a total of 577,889 bacterial sequences (mean± standard error/sample 1858±11, *n*=311), comprising of 240 genera and 444 distinct OTUs were classified to species level (Table A6.1).

6.3.1 Partitioning of species

A DAR was used to partition species into common, most persistent and abundant, and rare, transient taxa found in low abundance. The log abundance of each OTU was plotted against persistence (the number of samples in which a particular species appeared) showing a significant positive relationship ($F_{(1,433)}$=29.32, P<0.001), presented in Figure 6.1. Species present in more than 75% of the samples were considered to be common, while the remaining species were rare.

A total of five OTUs representing distinct species were considered to be common when the whole data set was considered; *Prevotella melaninogenica, Rothia mucilaginosa, Streptococcus mitis, Veillonella parvula, Pseudomonas aeruginosa*. These OTUs were found to make up over 55% of the total sequences in the study (318,866 sequences), while the rare OTUs made up just over 44% of the total sequences (259,023 sequences).

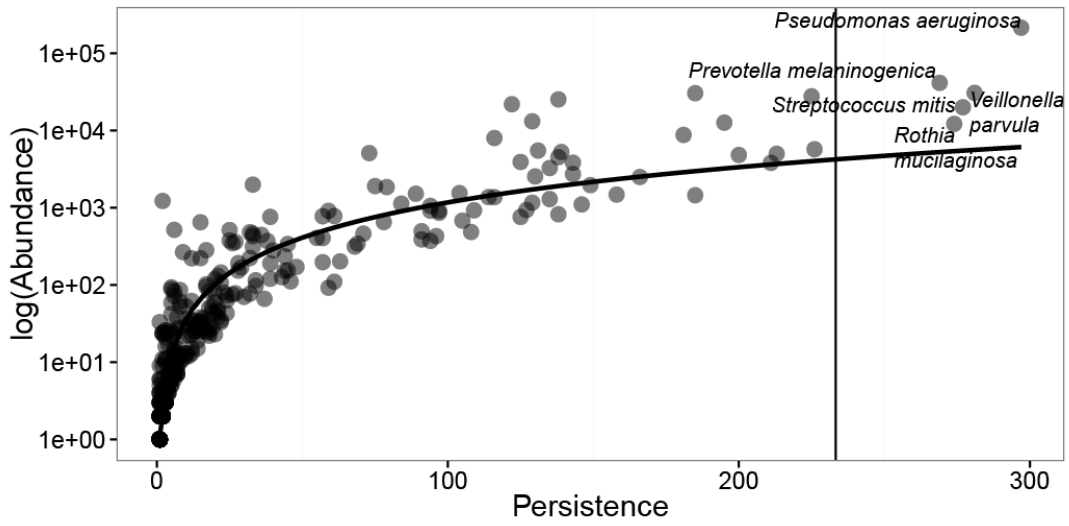


Figure 6.1 Persistence abundance curve of all OTUs.

All OTUS were plotted with Persistance, the number of samples in which they appear, against log abundance, total number of sequences, in order to partition OTUs into those that are common and those that are rare. Common OTUs were defined as those that fell in the upper quartile (right of the vertical line), all OTUs that fell below the line were considered to be rare (n=311, $F_{(1,433)}$=29.32, P<0.001).

6.3.2 Diversity

Three recognised measures of community diversity were applied to the data, species richness (*S**), Shannon-Wiener index (*H'*), and inverse Simpson's index (1-*D*). As varying levels of sampling depth in this study were accounted for during the sequencing analysis step, resampling, as performed in previous chapters, was not carried out on these data. Bacterial diversity was found to be highly variable between samples. In order to investigate if clinical factors were influencing this variation, linear relationships between bacterial diversity and clinical factors were investigated, these are highlighted below.

6.3.2.1 Lung function (FEV_1)

Published data suggests that a significantly positive linear relationship exists between diversity and lung function. To investigate if this result was not purely an artefact of the small sample sets used, the relationship between bacterial diversity (as measured by *S**, *H'* and 1-*D*) was modelled against lung function. Lung function data was available from 95% of the sequenced sputum samples. As shown in Figure 6.2, a significant positive relationship was seen between lung function, forced expiratory volume in 1 second (FEV_1), and the 3 measures of diversity investigated. Model results are show in Table 6.3.

6.3.2.2 Age

It was hypothesised that this relationship would be mirrored with patient age. Patient age at the time of sample collection was available for 87% of the CF samples sequenced. However no significant relationship was found between diversity and patient age for any of the diversity measures investigated, as shown in Table 6.3.

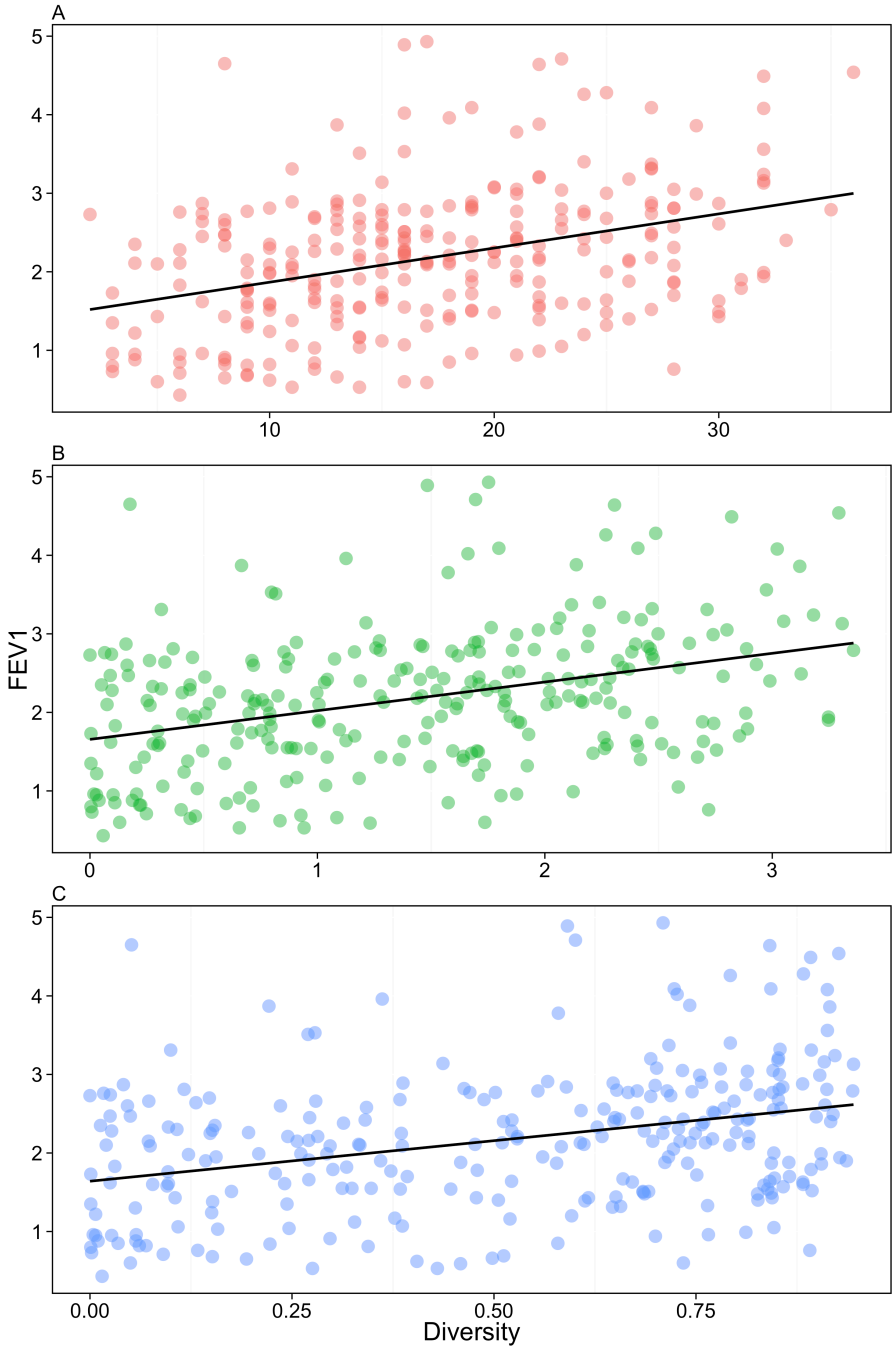


Figure 6.2 The relationship between diversity and lung function (FEV1).

A significant positive linear relationship was observed using a linear regression model, fitted to the data for each of the following measures of diversity; A) Species richness (P<0.001, r^2=0.22, $F_{(1,293)}$=82.8) B) Shannon-Wiener index (P<0.001, r^2=0.21, $F_{(1,293)}$=79.92) C) Inverse Simpson's index (P<0.001, r^2=0.17, $F_{(1,293)}$=61.95).

Table 6.3 Summary of linear regression statistics between bacterial diversity and continuous clinical factors.

Linear regressions were calculated to predict the relationship between bacterial diversity, species richness (*S*),* Shannon-weiner (*H')* and Inverse Simpson's (1-*D*), and continuous clinical factors, forced expiratory volume in 1 second (FEV_1), Age and body mass index (BMI). Significant relationships (*P*-value <0.001) are indicated in bold. *n*; number of samples included in each analysis, df; degrees of freedom, *F*; *F*-statistic describes the explained variation over the unexplained varation, *r*2; the proportion of variation in the y-variable due to variation in the x-variable

Clinical factor		Diversity	*n*	df	*F*	r^2	*P*
FEV_1	All	S*	276	1,274	43.11	0.136	**<0.001**
		H'	276	1,274	41.8	0.132	**<0.001**
		1-D	276	1,274	36.53	0.12	**<0.001**
	Common	S*	276	1,274	43.11	0.133	**<0.001**
		H'	276	1,274	37.11	0.119	**<0.001**
		1-D	276	1,274	29.51	0.097	**<0.001**
	Rare	S*	276	1,274	35.03	0.113	**<0.001**
		H'	276	1,274	11.28	0.04	**<0.001**
		1-D	276	1,274	9.32	0.034	**0.002**
Age	All	S*	254	1,252	0.006	<0.001	0.94
		H'	254	1,252	<0.001	<0.001	0.998
		1-D	254	1,252	0.01	<0.001	0.92
	Common	S*	254	1,252	0.006	<0.001	0.939
		H'	254	1,252	0.524	0.002	0.47
		1-D	254	1,252	0.805	0.003	0.37
	Rare	S*	254	1,252	0.017	<0.001	0.897
		H'	254	1,252	2.959	0.012	0.086
		1-D	254	1,252	2.458	0.01	0.118
BMI	All	S*	183	1,181	7.252	0.039	**0.008**
		H'	183	1,181	4.583	0.02	**0.03**
		1-D	183	1,181	3.002	0.02	0.08
	Common	S*	183	1,181	7.252	0.039	**0.008**
		H'	183	1,181	5.278	0.028	**0.023**
		1-D	183	1,181	5.196	0.028	**0.02**
	Rare	S*	183	1,181	8.517	0.045	**0.004**
		H'	183	1,181	1.72	0.009	0.191
		1-D	183	1,181	0.866	0.005	0.353

6.3.2.3 Body mass index

The relationship between body mass index (BMI) and bacterial community diversity was investigated using linear regression. BMI at the time of sample collection was recorded from 63% of patients and was modelled against each of the three diversity measures. As shown in Table 6.3, using analysis of **variance**, a significant relationship (*P*-value <0.001) was seen between BMI and species richness and the Shannon-Wiener index however, no significant relationship was observed between BMI and the Inverse Simpson's index.

6.3.3 Analysis of variance (ANOVA)

To investigate relationships between bacterial diversity and categorical clinical factors ANOVAs were employed, results are shown in Table 6.4. Bacterial diversity was not found to be related to patient gender, liver disease, pancreatic insufficiency, diabetes or CFTR genotype. Significant differences in diversity were only observed as a result of clinical status (stable, exacerbation or control) and location.

Clinical status was found to be significant for the whole community as well as the partitioned common and rare groups. Post-hoc Tukey HSD test revealed that the bacterial diversity of control samples was significantly different from the diversity observed between stable and exacerbating patients (*P*<0.001) as illustrated in Figure 6.3. While the variability in diversity in both CF groups was found to be high, the change in diversity between stable and exacerbating patients was also significant (*P*=0.009).

Table 6.4 Summary statistics of ANOVA comparing bacterial diversity to categorical clinical data.

Analysis of the changes in diversity, species richness (*S*),* Shannon-weiner (*H')* and Inverse Simpson's (1-*D*) of the whole bacterial community, common and rare taxa groups associated with clinical factors, including; gender, diabetes and liver disease. P- values of <0.05 are considered significant and shown in bold.

Clinical factor		Diversity	*n*	df	*F*	r^2	*P*
Gender	All	*S**	292	1,290	31.18	0.01	0.078
		H'	292	1,290	2.871	0.01	0.091
		1-*D*	292	1,290	43.652	0.02	**0.032**
	Common	*S**	292	1,290	3.118	0.01	0.079
		H'	292	1,290	1.493	0.005	0.223
		1-*D*	292	1,290	0.706	0.002	0.402
	Rare	*S**	292	1,290	0.372	0.001	0.542
		H'	292	1,290	0.102	<0.001	0.749
		1-*D*	292	1,290	0.266	<0.001	0.606
Diabetes	All	*S**	287	1,285	0.584	0.02	0.445
		H'	287	1,285	0.268	<0.001	0.605
		1-*D*	287	1,285	0.908	0.005	0.908
	Common	*S**	287	1,285	0.584	0.002	0.445
		H'	287	1,285	0.464	0.002	0.496
		1-*D*	287	1,285	0.173	<0.001	0.678
	Rare	*S**	287	1,285	2.668	0.009	0.104
		H'	287	1,285	0.131	<0.001	0.717
		1-*D*	225	1,285	0.047	<0.001	0.829
Liver disease	All	*S**	225	1,223	0.606	0.003	0.437
		H'	225	1,223	0.226	0.001	0.635
		1-*D*	225	1,223	0.357	0.005	0.551
	Common	*S**	225	1,223	0.606	0.003	0.437
		H'	225	1,223	0.031	<0.001	0.861
		1-*D*	225	1,223	0.241	0.112	0.624
	Rare	*S**	225	1,223	0.045	<0.001	0.832
		H'	225	1,223	0.56	0.003	0.455
		1-*D*	225	1,223	0.373	0.002	0.542

Table 6.4 Continued

Clinical factor		Diversity	*n*	df	*F*	r^2	*P*
Pancreatic insufficiency	All	*S**	225	1,223	0.467	0.002	0.495
		H′	225	1,223	0.13	<0.001	0.709
		1-*D*	225	1,223	0.835	0.003	0.362
	Common	*S**	225	1,223	0.467	0.002	0.495
		H′	225	1,223	1.012	0.004	0.315
		1-*D*	225	1,223	0.915	0.004	0.34
	Rare	*S**	225	1,223	2.233	0.01	0.137
		H′	225	1,223	0.455	0.002	0.501
		1-*D*	225	1,223	0.131	<0.001	0.718
Clinical status	All	*S**	269	1,232	31.53	0.204	**<0.001**
		H′	269	1,232	32.01	0.216	**<0.001**
		1-*D*	269	1,232	20.18	0.148	**<0.001**
	Common	*S**	269	1,232	45.69	0.164	**<0.001**
		H′	269	1,232	8.826	0.037	**0.003**
		1-*D*	269	1,232	5.265	0.022	**0.023**
	Rare	*S**	269	1,232	52.51	0.312	**<0.001**
		H′	269	1,232	28.72	0.199	**<0.001**
		1-*D*	269	1,232	13.44	0.104	**<0.001**
Genotype	All	*S**	254	1,249	1.045	0.008	0.353
		H′	254	1,249	1.144	0.009	0.32
		1-*D*	254	1,249	1.125	0.009	0.326
	Common	*S**	254	1,249	1.045	0.008	0.353
		H′	254	1,249	0.356	0.003	0.701
		1-*D*	254	1,249	0.493	0.004	0.59
	Rare	*S**	254	1,249	0.763	0.006	0.218
		H′	254	1,249	0.191	0.002	0.558
		1-*D*	254	1,249	0.012	<0.001	0.989
Location	All	*S**	292	1,281	4.012	0.12	**<0.001**
		H′	292	1,281	3.371	0.107	**<0.001**
		1-*D*	292	1,281	3.005	0.097	**<0.001**
	Common	*S**	292	1,281	4.012	0.125	**<0.001**
		H′	292	1,281	2.36	0.077	**0.011**
		1-*D*	292	1,281	2.14	0.071	**0.022**
	Rare	*S**	292	1,281	4.325	0.133	**<0.001**
		H′	292	1,281	1.685	0.057	0.084
		1-*D*	292	1,281	1.098	0.037	0.364

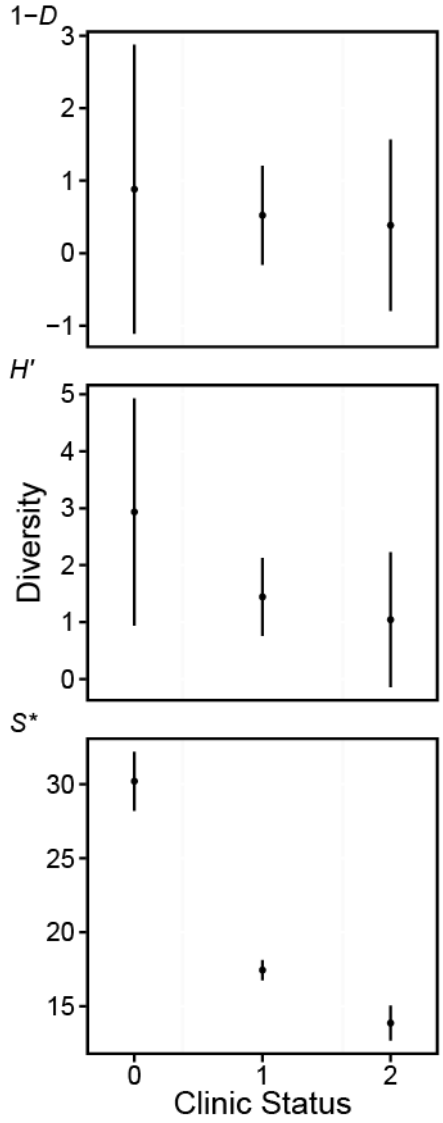


Figure 6.3 Changes in diversity associated with clinical status.

Each point represents the mean value of diversity for each clinical state, lines indicate the standard error of the mean. Clinical status 0; Non CF healthy controls, 1; Stable, 2; Cystic fibrosis pulmonary exacerbation (CFPE). Results of the associated ANOVA are shown in Table 6.4.

The relationship of bacterial diversity with geographical location was investigated using ANOVA, considering each unique site as a different variable. This analysis therefore did not account for the differences in distance between sites. As shown in Table 6.4, bacterial diversity was found to be significantly different between sample collection sites. These differences in bacterial diversity between sites are visualised in Figure 6.4, and show high variability between sites. A post-hoc Tukey HSD test was then used to assess the statistical differences between each centre. This analysis revealed that while the overall difference by location was significant, the difference in diversity between most centres was not significant.

Difference in species richness between, Southampton, UK and Belfast, UK (P=0.02), London, UK and Belfast, UK (P=0.005), Portland, ME and London, UK (P=0.04), Boston, MA and London, UK (P=0.02) were shown to be significant. When considering Shannon's diversity index, only Southampton, UK and Belfast, UK (P=0.03) and Belfast, UK and Vermont, ME (P=0.02) were found to be significantly different. Further, the significant difference observed between and Belfast, UK and Vermont, ME (P=0.03) observed using Shannon's diversity index was also observed using the inverse Simpson's index.

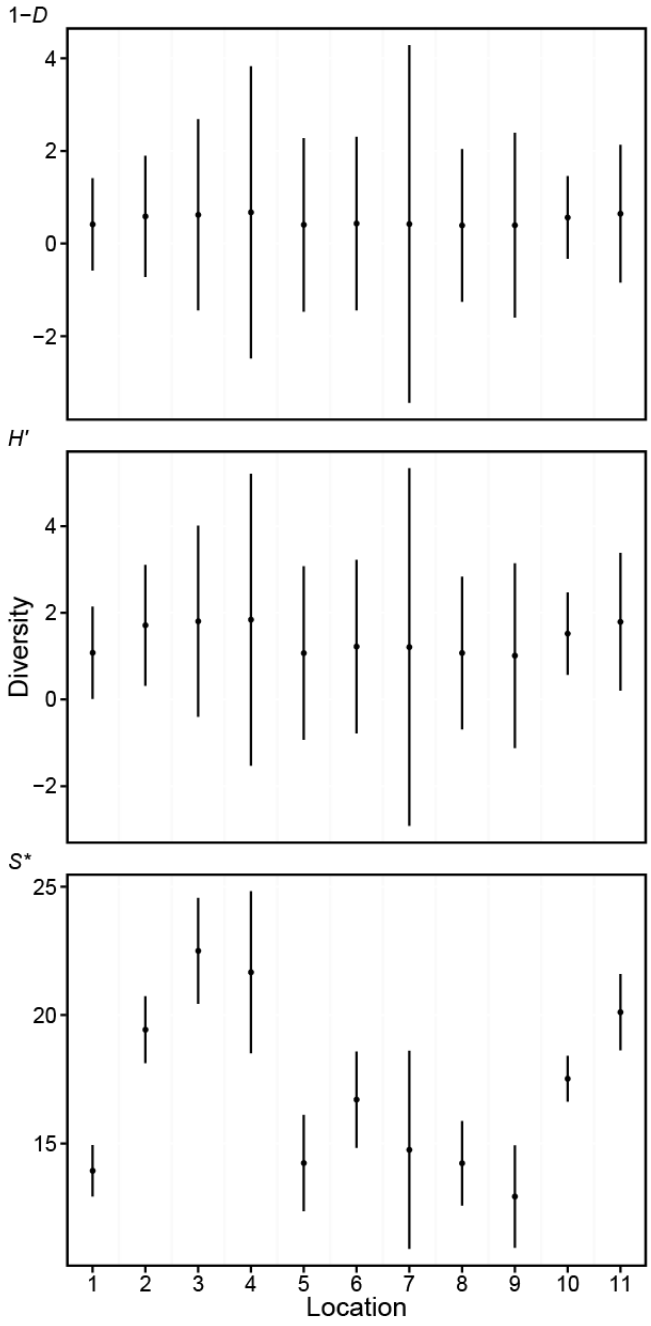


Figure 6.4 Changes in diversity associated with Location.

Each point represents the mean value of diversity for each location, lines indicate the standard error of the mean. Location 1; Belfast, UK, 2; Southampton, UK, 3; London, UK, 4; Dublin, Eire, 5; Warsaw, Poland, 6; Dartmouth, Bedford, NH, 7; Dartmouth, Lebanon, NH, 8; Portland, ME, 9; Boston, MA, 10, Seattle, WA, 11; Vermont, ME. Results of the associated ANOVA are shown in Table 5.4.

6.3.4 Mantel and partial Mantel tests

To examine which clinical factors were related to community similarity, lower tailed Mantel and partial Mantel tests were employed. These tests compared the Bray-Curtis similarity of the whole community, as well as the common and rare groups, to distance matrices created for each clinical factor. It was hypothesised that patients with more similar communities would have more similar clinical outcomes and therefore a lower tailed Mantel test was used for every comparison. All clinical factors were compared to each other using a two-tailed Mantel test, this analysis allowed us to account for auto-correlation in the partial Mantel tests (Table A6.2). Factors shown to be auto-correlated were then examined using the lower-tailed test, with autocorrelation accounted for in subsequent analyses.

6.3.4.1 Mantel tests

Continuous clinical variables were compared to the similarities of the whole, common and rare taxa groups. It was hypothesised that patients with more similar lung function (FEV_1), age and BMI would have more similar bacterial communities. This relationship was confirmed by the Mantel test (Table 6.5) which showed a significant negative relationship between community similarity and all continuous clinical factors ($P<0.001$).

Categorical clinical variables were also compared to community similarity using Mantel test, shown in Table 6.5. In order not to weight the categories, matrices were calculated so that samples falling into the same categories were coded as 0 and those that were different were coded as 1. Genotype, was found to show no significant relationship with community composition (whole; $P=0.92$, common; $P=0.972$, rare; $P=0.282$). Table 6.5 also shows that diabetes, gender and pancreatic insufficiency were found to only relate to significant differences in community composition when compared to the rare taxa group. Conversely the relationship between community composition and both liver disease and clinical status (whether a patient is considered clinically stable or experiencing a CF pulmonary exacerbation (CFPE)) were found to be significant in all groups (Table 6.5).

Table 6.5 Mantel test analyses for the association between bacterial community composition and clinical factors.

The *P*-values have been calculated using the distribution of the Mantel test statistic (*r*) estimated for 9999 permutations. *N*= number of samples, *n*= number of pair wise comparisons between groups. Significant *P*-values are indicated in bold.

Clinical Factor	*N*	*n*		*r*	*P*
Location	292	42486	All	-0.024	**<0.001**
			Common	-0.022	**<0.001**
			Rare	-0.044	**<0.001**
FEV_1	276	37950	All	-0.098	**<0.001**
			Common	-0.106	**<0.001**
			Rare	-0.035	**<0.001**
Clinical status	217	23436	All	-0.031	**<0.001**
			Common	-0.04	**<0.001**
			Rare	-0.069	**<0.001**
Age	255	32385	All	-0.022	**<0.001**
			Common	-0.021	**<0.001**
			Rare	-0.015	**<0.001**
BMI	182	16471	All	-0.066	**<0.001**
			Common	-0.07	**<0.001**
			Rare	0.028	1
Diabetes	287	41041	All	-0.003	0.318
			Common	0.001	0.724
			Rare	-0.032	**<0.001**
Gender	292	42486	All	0.001	0.575
			Common	0.001	0.579
			Rare	-0.022	**<0.001**
Genotype	252	31626	All	0.006	0.92
			Common	0.008	0.972
			Rare	-0.002	0.282
Liver Disease	226	25425	All	-0.094	**<0.001**
			Common	-0.092	**<0.001**
			Rare	-0.041	**<0.001**
Pancreatic insufficiency	225	25200	All	0.029	1
			Common	0.03	1
			Rare	-0.013	**<0.001**

To examine how geographical location affected the bacterial community composition samples were partitioned by centre. The data was split into 11 geographical locations, shown in Table 6.1. Coordinates, shown in Table 6.6, from each of the clinics who provided samples were collected and the distances between centres were calculated. It was hypothesised that the bacterial community composition would be more similar in clinics closer in geographical distance whereas those further apart would be more distinct. Mantel tests were used to investigate this relationship, revealing a significant correlation between community similarity and geographical distance (P<0.001). This result was observed for the whole community and the common and rare groups, Table 6.5.

6.3.4.2 Partial Mantel tests

Partial Mantel test were used to further investigate the observed relationship between FEV_1 and bacterial community composition. FEV_1 was compared to community composition for the whole community as well as the common and rare groups, controlling for each clinical factor in turn; shown in Table 6.7. As a result of some values being unavailable for all clinical factors a subset of the data was investigated using partial Mantel test. This subset consisted of 79% of the total samples, from 9 of the original 11 centres. Consequently, all samples from Belfast and Seattle were removed due to a lack of associated clinical metadata.

This analysis revealed that for the whole community and common group, only clinical status (P<0.001), BMI (P<0.001) and liver disease (P<0.001) were correlated with FEV_1 and bacterial community composition (Table 6.7). However, the rare OTU group was found to correlate with location (P<0.001), clinical status (P=0.008), gender (P<0.001) and pancreatic insufficiency (P<0.001).

Table 6.6 List of clinic locations and co-ordinates.

Exact clinic locations were not available for samples collected in UK, Eire or Poland. Therefore co-ordinates relate to central locations in each of the county capitals.

Reference	Location	Clinic	Coordinates
1	*Belfast, UK*	Belfast city hospital	54.5875° N, 5.9408° W
2	*Southampton, UK*	Southampton general hospital	50.9330° N, 1.4340° W
3	*London, UK*	Belfast city hospital	51.5000° N, 0.1167° W
4	*Dublin, Eire*	Belfast city hospital	53.0000° N, 8.0000° W
5	*Warsaw, Poland*	Belfast city hospital	52.2167° N, 21.0333° E
6	*Dartmouth, Bedford, NH*	Children's Hospital at Dartmouth-Hitchcock	43.6761° N, 71.4228822° W
7	*Dartmouth, Lebanon, NH*	Dartmouth–Hitchcock Medical Center	43.6761° N, 72.2733° W
8	*Portland, ME*	Maine Medical Center	43.6530° N, 70.2760° W
9	*Boston, MA*	Massachusetts General Hospital	42.3628° N, 71.0686° W
10	*Seattle, WA*	Seattle general hospital	47.6628° N, 122.2817° W
11	*Vermont, ME*	Vermont Medical centre	44.479794°N 73.194119°W

Table 6.7 Partial Mantel test analyses for the association between bacterial community composition and clinical factors.

A) The correlation with the whole community, B) the correlation with the common OTU group and C) the correlation with the rare OTU group. The *P*-values have been calculated using the distribution of the Mantel test statistic (*r*) estimated for 9999 permutations. *N*= number of samples, *n*= number of pair wise comparisons between groups. Significant *P*-values are indicated in bold.

A

Clinical factor	Control for	*r*	*P*	Clinical factor	Control for	*r*	*P*
	Location	**-0.145**	***<0.001***	Location		0.016	0.961
	status	-0.145	**<0.001**	clinical status		-0.075	**<0.001**
	Age	-0.146	**<0.001**	Age		0.025	0.996
	BMI	-0.145	**<0.001**	BMI		-0.038	**<0.001**
FEV1	Diabetes	-0.146	**<0.001**	Diabetes	FEV1	0.007	0.771
	Gender	-0.146	**<0.001**	Gender		-0.004	0.33
	Liver disease	-0.14	**<0.001**	Liver disease		-0.109	**<0.001**
	Pancreatic insufficiency	-0.146	**<0.001**	Pancreatic insufficiency		-0.004	0.315

B

Clinical factor	Control for	*r*	*P*	Clinical factor	Control for	*r*	*P*
	Location	**-0.15**	***<0.001***	Location		0.019	0.98
	status	-0.15	**<0.001**	clinical status		-0.085	**<0.001**
	Age	-0.15	**<0.001**	Age		0.015	0.952
	BMI	-0.149	**<0.001**	BMI		-0.05	**<0.001**
FEV1	Diabetes	-0.15	**<0.001**	Diabetes	FEV1	0.007	0.778
	Gender	-0.15	**<0.001**	Gender		0.001	0.555
	Liver disease	-0.144	**<0.001**	Liver disease		-0.107	**<0.001**
	Pancreatic insufficiency	-0.15	**<0.001**	Pancreatic insufficiency		0.004	0.669

Table 6.7 Continued

C

Clinical factor	Control for	*r*	*P*	Clinical factor	Control for	*r*	*P*
FEV1	Location	**-0.055**	***<0.001***	Location	FEV1	-0.104	**<0.001**
	status	-0.053	**<0.001**	clinical status		-0.022	**0.008**
	Age	-0.053	**<0.001**	Age		-0.001	0.45
	BMI	-0.054	**<0.001**	BMI		0.028	0.999
	Diabetes	-0.053	**<0.001**	Diabetes		-0.012	0.104
	Gender	-0.053	**<0.001**	Gender		-0.041	**<0.001**
	Liver disease	-0.053	**<0.001**	Liver disease		-0.009	0.17
	Pancreatic insufficiency	-0.053	**<0.001**	Pancreatic insufficiency		-0.036	**<0.001**

6.3.5 Analysis of similarity (ANOSIM)

Using both Sørensen and Bray-Curtis similarity, ANOSIM's were carried out to further investigate the relationship between genotype and bacterial community composition. No significant difference was observed using Sørensen similarity index, however results comparing Bray Curtis similarity showed a significant difference in the whole bacterial community when homogeneous ΔF508 genotype was compared to heterogeneous ΔF508 genotype (*P*=0.0379) (Table A6.3).

Using Mantel tests it was revealed that when comparing genders, there was a significant change in the community composition of the rare OTU group. ANOSIMs, using both Sørensen and Bray-Curtis similarity, were used to further investigate this relationship. No significant difference was observed between males and females using Sørensen for the whole community (*R*=0.008, *P*=0.202) or the common OTU group (*R*=0.002, *P*=0.403) however, a significant difference was observed between genders in the rare OTU group (*R*=0.029, *P*=0.021). When using Bray-Curtis similarity index, a significant difference was observed in the whole community (*R*=0.026, *P*=0.029). From partitioning it was clear that this difference in community composition was driven by the rare taxa (*R*=0.032, *P*=0.01) as no significant difference was observed in the common OTU group (*R*=0.003, *P*=0.331).

Mantel tests revealed that the community composition from samples collected from sites closer in geographic distance were more similar than those from geographically distinct locations. Analysis of similarity (ANOSIM) was used to back up these results and determine how the community composition changed between centre using both the Sørensen (Table 6.8) and Bray-Curtis (Table 6.9) measures of similarity.

The non CF healthy control group was included in this analysis as a distinct location. Using Sørensen similarity coefficient, the control group was found to be significantly different from all centres with the exception samples from of Southampton (*R*=0.049, *P*=0.161). However, when examining changes in OTU presence and abundance using Bray-Curtis similarity, the control samples were found to be significantly different from all centres. While Sørensen similarity showed the common taxa group to be similar to all centres apart from Poland

(*R*=0.075, *P*=0.009) and Bedford, Dartmouth (*R*=0.073, *P*=0.009), all centres were found to be significantly different when accounting for species abundance using Bray-Curtis similarity measure. When the rare group was explored, similar community composition was observed with both Sørensen and Bray-Curtis similarity measures, revealing similar composition in the control samples as those from Belfast (Sørensen; *R*=-0.045, *P*=0.711, Bray-Curtis, *R*=-0.0143, *P*=0.992), Southampton (Sørensen; *R*=-0.0131, *P*=0.552, Bray-Curtis, *R*=-0.0088, *P*=0.514) and Seattle (Sørensen; *R*=0.059, *P*=0.206, Bray-Curtis, *R*=-0.007, *P*=0.517).

ANOSIMs performed using Sørensen measure revealed the common taxa to be conserved between groups, with only samples from Seattle, WA (e.g. Belfast, UK; *R*=0.042, *P*=0.007) and Vermont, ME (eg. Portland, ME; *R*=0.109, *P*=0.004) showing significant differences from other clinic locations, Table 6.8B. On the other hand, significant differences in rare taxa were found to occur when compared to least one other CF centre, Table 6.8C. Belfast, UK, was the only centre to be found to be significantly different from a single other location, in this case Seattle, WA (*R*=0.134, *P*<0.001). Comparison of the differences in whole community composition and the common and rare groups indicated that the rare community was driving differences between centres.

A similar pattern was observed when comparing community composition between CF centres using Bray-Curtis measure of similarity. However, when considering OTU abundance, more significant differences were observed in the common OTU group. This was particularly prevalent between samples from Belfast, UK, which were shown to have significantly different community composition from 4 other sites (Dartmouth, Bedford, NH, Portland, ME, Seattle, WA and Vermont, ME) and Vermont, ME, which was shown to be significantly different from 6 other sites (Belfast, UK, Warsaw, Poland, Dartmouth, Lebanon, NH, Portland, ME, Boston, MA and Seattle, WA), Table 6.9B. When the rare OTU group was compared more significant differences were observed, again suggesting that the rare community is responsible for driving the changes in community composition between CF centres.

Table 6.8 Comparison of community composition using analysis of similarity (ANOSIM) for each centre location.

ANOSIMs were carried out using Sorensen measure of similarity, for the A) whole, B) common and C) rare communities. Matrices below show *R* values in the lower half and *P* values on the upper. Significant *R* values indicate the samples within a group are more similar than would be expected by random chance, therefore the two groups can be considered significantly different. 1) *Belfast, UK,* 2) Southampton, UK, 3) *London, UK*, 4) *Dublin, Eire*, 5) *Warsaw, Poland*, 6) *Dartmouth, Bedford, NH*, 7) *Dartmouth, Lebanon, NH,* 8) *Portland, ME*, 9) *Boston, MA,* 10) *Seattle, WA*, 11) *Vermont, ME*.

P values less than 0.05 were considered significant. Significant *P* values are highlight in green.

A

	1	2	3	4	5	6	7	8	9	10	11	12
1	r^2 *P*	0.22	0.8542	0.3681	0.6404	0.3345	0.444	0.0995	0.0703	0.0001	0.6653	0.0473
2	0.0261	r^2 *P*	0.9242	0.4748	0.1288	0.0648	0.2457	0.0006	0.007	0.0008	0.0563	0.1605
3	-0.087	-0.097	r^2 *P*	0.0001	0.0092	0.0003	0.005	0.0008	0.0003	0.1137	0.0392	0.0001
4	0.031	-0.002	0.535	r^2 *P*	0.1721	0.0044	0.0049	0.0352	0.2279	0.0453	0.0382	0.0002
5	-0.031	0.0693	0.139	0.1057	r^2 *P*	0.0045	0.1397	0.0003	0.0199	0.0003	0.0023	0.0001
6	0.0254	0.0887	0.3126	0.3603	0.1526	r^2 *P*	0.6166	0.7063	0.0648	0.0005	0.1968	0.0001
7	0.0025	0.0966	0.595	0.7381	0.1727	-0.047	r^2 *P*	0.5933	0.9171	0.0416	0.2987	0.0016
8	0.0814	0.1795	0.2957	0.2301	0.2075	-0.025	-0.052	r^2 *P*	0.0734	0.0002	0.0293	0.0001
9	0.1238	0.1952	0.2354	0.0834	0.1083	0.0679	-0.208	0.0764	r^2 *P*	0.0002	0.0064	0.0001
10	0.1338	0.164	0.0972	0.2094	0.2711	0.2824	0.2817	0.2738	0.3703	r^2 *P*	0.0326	0.0024
11	-0.024	0.048	0.1286	0.2427	0.1904	0.0339	0.0695	0.0704	0.2044	0.1078	r^2 *P*	0.0001
12	0.1208	0.0496	0.5099	0.8882	0.548	0.5801	0.6969	0.5013	0.6073	0.2227	0.3464	r^2 *P*

B

	1	2	3	4	5	6	7	8	9	10	11	12
1	r^2 *P*	0.9215	0.887	0.8222	0.6527	0.5663	0.3859	0.0312	0.1369	0.0073	0.8519	0.9926
2	-0.053	r^2 *P*	0.6037	0.677	0.2041	0.1365	0.1409	0.0009	0.017	0.2836	0.2853	0.9174
3	-0.108	-0.025	r^2 *P*	0.4542	0.2532	0.336	0.0678	0.9069	0.1029	0.4902	0.654	0.0642
4	-0.154	-0.083	-4E-04	r^2 *P*	0.8587	0.7934	0.2364	0.9854	0.9845	0.5141	0.596	0.4344
5	-0.037	0.0533	0.0095	-0.14	r^2 *P*	0.9655	0.3601	0.7324	0.1333	0.0631	0.1787	0.0099
6	-0.021	0.076	0.0053	-0.142	-0.025	r^2 *P*	0.4158	0.9269	0.18	0.0484	0.1954	0.0099
7	0.0343	0.1966	0.3141	0.1548	0.0628	0.0269	r^2 *P*	0.8198	0.6739	0.0957	0.1532	0.0673
8	0.1421	0.1885	-0.064	-0.236	-0.025	-0.04	-0.154	r^2 *P*	0.6523	0.0001	0.0037	0.0386
9	0.1024	0.192	0.0311	-0.221	0.0278	0.019	-0.099	-0.026	r^2 *P*	0.0018	0.0241	0.0023
10	0.0419	0.0232	-0.006	-0.048	0.1314	0.1462	0.2639	0.3568	0.3117	r^2 *P*	0.2731	0.9171
11	-0.058	0.009	-0.035	-0.081	0.0448	0.042	0.2031	0.1089	0.1536	0.031	r^2 *P*	0.9621
12	-0.155	-0.063	0.0529	0.0739	0.0748	0.0729	0.373	0.0525	0.1532	-0.1	-0.046	r^2 *P*

Table 6.8 Continued.

C. Comparison of rare community composition using analysis of similarity (ANOSIM) for each centre location.

1) *Belfast, UK,* 2) Southampton, UK, 3) *London, UK*, 4) *Dublin, Eire*, 5) *Warsaw, Poland*, 6) *Dartmouth, Bedford, NH*, 7) *Dartmouth, Lebanon, NH,* 8) *Portland, ME*, 9) *Boston, MA,* 10) *Seattle, WA*, 11) *Vermont, ME*.

	1	*2*	*3*	*4*	*5*	*6*	*7*	8	*9*	*10*	*11*	*12*
1	r^2 *P*	0.6024	0.9939	0.6388	0.5378	0.5001	0.609	0.2607	0.0819	0.0001	0.8722	0.7112
2	-0.014	r^2 *P*	0.9478	0.487	0.0683	0.0516	0.2564	0.0006	0.0056	0.0082	0.0344	0.5518
3	-0.185	-0.118	r^2 *P*	0.0001	0.0225	0.0005	0.0041	0.0006	0.0008	0.3873	0.0839	0.0001
4	-0.057	-0.013	0.5505	r^2 *P*	0.2948	0.0121	0.005	0.0196	0.2963	0.1453	0.068	0.0002
5	-0.012	0.1028	0.1016	0.0629	r^2 *P*	0.0008	0.2337	0.0002	0.0134	0.0009	0.0017	0.0001
6	-0.005	0.1081	0.2854	0.3212	0.1739	r^2 *P*	0.7071	0.6067	0.046	0.0015	0.1786	0.0001
7	-0.063	0.1101	0.6116	0.7857	0.123	-0.086	r^2 *P*	0.4731	0.9549	0.062	0.3559	0.0016
8	0.0349	0.1939	0.3102	0.2664	0.2642	-0.015	-0.018	r^2 *P*	0.0186	0.001	0.0299	0.0001
9	0.1178	0.2233	0.2102	0.0502	0.1184	0.0772	-0.235	0.1266	r^2 *P*	0.0001	0.0029	0.0001
10	0.1342	0.1235	0.0198	0.1395	0.2807	0.2653	0.261	0.2463	0.3729	r^2 *P*	0.0576	0.2062
11	-0.059	0.0539	0.0984	0.2087	0.2078	0.0398	0.0494	0.0685	0.2202	0.0945	r^2 *P*	0.0003
12	-0.045	-0.013	0.4979	0.9086	0.4836	0.5331	0.7048	0.4711	0.5514	0.0595	0.2746	r^2 *P*

Table 6.9 Comparison of community composition using analysis of similarity (ANOSIM) for each centre location.

ANOSIMs were carried out using Bray-Curtis measure of similarity, for the A) whole, B) common and C) rare communities. Matrices below show *R* values in the lower half and *P* values on the upper. Significant *R* values indicate the samples within a group are more similar than would be expected by random chance, therefore the two groups can be considered significantly different. 1) *Belfast, UK,* 2) Southampton, UK, 3) *London, UK*, 4) *Dublin, Eire*, 5) *Warsaw, Poland*, 6) *Dartmouth, Bedford, NH*, 7) *Dartmouth, Lebanon, NH,* 8) *Portland, ME*, 9) *Boston, MA,* 10) *Seattle, WA*, 11) *Vermont, ME*.

P values less than 0.05 were considered significant. Significant *P* values are highlighted in green.

A

	1	2	3	4	5	6	7	8	9	10	11	12
1	r^2 P	0.0003	0.0295	0.7071	0.977	0.0305	0.0919	0.0295	0.173	0.0002	0.0016	0.0001
2	0.169	r^2 P	0.7363	0.8735	0.3203	0.0576	0.1413	0.0195	0.133	0.0527	0.2608	0.0122
3	0.1763	-0.044	r^2 P	0.0649	0.001	0.3053	0.0308	0.0325	0.0267	0.0611	0.201	0.0001
4	-0.087	-0.135	0.1615	r^2 P	0.0159	0.7382	0.0047	0.6338	0.936	0.7903	0.5422	0.0001
5	-0.136	0.0207	0.2943	0.365	r^2 P	0.0108	0.0193	0.0721	0.024	0.6996	0.0118	0.0001
6	0.1574	0.0847	0.0109	-0.085	0.116	r^2 P	0.3516	0.7966	0.37	0.0025	0.1156	0.0001
7	0.2316	0.1644	0.2386	0.5159	0.4841	0.036	r^2 P	0.3307	0.3389	0.0273	0.0403	0.001
8	0.1319	0.0953	0.117	-0.062	0.0619	-0.032	0.0651	r^2 P	0.4035	0.0105	0.0898	0.0001
9	0.0765	0.0725	0.1197	-0.176	0.0881	0.0028	0.0565	0.0005	r^2 P	0.0276	0.0204	0.0001
10	0.108	0.0582	0.1021	-0.079	-0.031	0.1849	0.2605	0.1356	0.1324	r^2 P	0.1123	0.0002
11	0.1811	0.012	0.0405	-0.02	0.1391	0.0516	0.2478	0.0443	0.1267	0.0493	r^2 P	0.0008
12	0.4491	0.131	0.3909	0.8571	0.8583	0.4659	0.8103	0.4288	0.6368	0.1954	0.2139	r^2 P

B

	1	2	3	4	5	6	7	8	9	10	11	12
1	r^2 P	0.0012	0.0349	0.8561	0.9881	0.044	0.1651	0.0307	0.2119	0.0003	0.0021	0.0001
2	0.1439	r^2 P	0.3899	0.8974	0.0543	0.1522	0.1286	0.0217	0.1066	0.2422	0.313	0.0073
3	0.1717	0.0125	r^2 P	0.2129	0.0008	0.2324	0.1168	0.167	0.0526	0.0841	0.263	0.002
4	-0.152	-0.119	0.0656	r^2 P	0.0454	0.7379	0.0377	0.8074	0.9663	0.954	0.226	0.0001
5	-0.154	0.0885	0.3099	0.264	r^2 P	0.0088	0.0199	0.1955	0.0193	0.7474	0.0065	0.0001
6	0.1415	0.0433	0.022	-0.072	0.1487	r^2 P	0.5728	0.73	0.2858	0.0844	0.1052	0.0001
7	0.1732	0.1312	0.1602	0.3333	0.4677	-0.027	r^2 P	0.4662	0.3707	0.0914	0.0496	0.0007
8	0.1365	0.0904	0.0491	-0.122	0.0254	-0.027	-0.02	r^2 P	0.569	0.0178	0.0419	0.0001
9	0.0662	0.0718	0.0945	-0.188	0.1015	0.0105	0.0255	-0.018	r^2 P	0.0675	0.0104	0.0001
10	0.0806	0.0177	0.0766	-0.132	-0.036	0.0646	0.138	0.11	0.0916	r^2 P	0.4138	0.0008
11	0.1769	0.0071	0.0249	0.0558	0.1768	0.047	0.1811	0.0642	0.1469	0.0049	r^2 P	0.0085
12	0.4539	0.1416	0.2248	0.9312	0.8343	0.3571	0.73	0.3187	0.5677	0.1604	0.1291	r^2 P

Table 5.9 Continued.

1) *Belfast, UK,* 2) Southampton, UK, 3) *London, UK*, 4) *Dublin, Eire*, 5) *Warsaw, Poland*, 6) *Dartmouth, Bedford, NH*, 7) *Dartmouth, Lebanon, NH,* 8) *Portland, ME*, 9) *Boston, MA,* 10) *Seattle, WA*, 11) *Vermont, ME*.

C

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1	r^2 *P*	0.4969	0.9935	0.7131	0.6179	0.36	0.2517	0.1755	0.1199	0.0001	0.6784	0.992
2	-0.003	r^2 *P*	0.8774	0.2307	0.0313	0.0022	0.089	0.0003	0.0051	0.0418	0.0036	0.514
3	-0.176	-0.087	r^2 *P*	0.0005	0.5108	0.0012	0.0058	0.0014	0.0056	0.2154	0.0691	0.0001
4	-0.079	0.0918	0.4319	r^2 *P*	0.1258	0.0001	0.0032	0.0217	0.6996	0.031	0.1305	0.0001
5	-0.024	0.1201	-0.011	0.131	r^2 *P*	0.0027	0.0789	0.0003	0.0278	0.0002	0.0004	0.0001
6	0.0205	0.1963	0.2332	0.4437	0.1686	r^2 *P*	0.1543	0.9635	0.0279	0.0001	0.2081	0.0001
7	0.0973	0.249	0.4842	0.6746	0.1929	0.1607	r^2 *P*	0.1903	0.6506	0.0059	0.0689	0.0002
8	0.0512	0.2217	0.246	0.2672	0.2013	-0.054	0.1466	r^2 *P*	0.0217	0.0001	0.0945	0.0001
9	0.0916	0.2202	0.1173	-0.08	0.0899	0.0869	-0.084	0.118	r^2 *P*	0.0001	0.0014	0.0001
10	0.1699	0.0774	0.06	0.2312	0.3142	0.3368	0.4424	0.2976	0.3682	r^2 *P*	0.0028	0.517
11	-0.026	0.0951	0.1006	0.14	0.2325	0.0365	0.2583	0.0408	0.2223	0.1637	r^2 *P*	0.0007
12	-0.143	-0.009	0.4512	0.5974	0.5051	0.5467	0.8201	0.42	0.4652	-0.007	0.1888	r^2 *P*

6.3.6 Similarity percentage (SIMPER)

Similarity percentage analysis (SIMPER) was carried out in order to examine the taxa responsible for the changes in community composition between genders, observed using ANOSIM. Using Bray-Curtis measure of similarity SIMPER tables were created to examine the contribution of species from the whole community as well as the common and rare groups, between genders (Table 6.10). *Pseudomonas aeruginosa* was found to contribute the most to the whole community and common group similarity between genders. Of the rare group, *Staphylococcus aureus* was observed to have the biggest contribution to community similarity.

The mean abundance of the recognised CF pathogens; *Pseudomonas aeruginosa, Staphylococcus aureus, Burkholderia cepacia* complex and *Stenotrophomonas maltophilia*, were all found to be higher in female patients than males.

Table 6.10 Similarity of percentages (SIMPER) analysis of the bacterial dissimilarity (Bray-Curtis) of the bacterial community between genders.

Given is the mean % abundance of sequences for each OTU present in females and males. In addition, the average dissimilarity between genders is given. Percentage contribution is calculated from the mean contribution divided by the mean dissimilarity between genders. A) SIMPER analysis of the whole community, B) SIMPER analysis of the common OTU group C) SIMPER analysis of the rare OTU group. *Indicates organisms partitioned as common; organisms in bold are considered to be recognised CF pathogens.

A

Taxon	% Mean abundance		Average dissimilarity	Contribution %	Cumulative %
	Female	Male			
Pseudomonas aeruginosa*	801	686	22.47	30.37	30.37
*Prevotella melaninogenica**	101	150	5.169	6.987	37.36
Staphylococcus aureus	120	93.1	4.992	6.747	44.11
***Burkholderia cepacia* complex**	113	67.7	4.484	6.061	50.17
Prophorymonas catoniae	74.4	101	3.935	5.318	55.48
Haemophilus influenzae	53	90.3	3.572	4.827	60.31
*Veillonella parvula**	69.7	96.3	3.096	4.185	64.5
Stenotrophomonas maltophilia	64.8	30.9	2.462	3.328	67.82
*Streptococcus mitis **	60.5	54.2	2.28	3.082	70.91
Prevotella	39	44.4	2.048	2.768	73.67
*Rothia mucilaginosa**	31.6	41.4	1.513	2.045	75.72
Prevotella oris	39.8	19.7	1.359	1.837	77.55
Neissaria flavescens	20.3	28	1.212	1.639	79.19
Prevotella denticola	20.8	16.7	0.9569	1.293	80.49
Streptococcus parasanguinis	14.5	17.8	0.7596	1.027	81.51

Table 6.10 Continued
B

Taxon	% Mean abundance		Average dissimilarity	Contribution %	Cumulative %
	Female	Male			
Pseudomonas aeruginosa*	801	686	40.26	59.45	59.45
*Prevotella melaninogenica**	101	150	10.99	16.23	75.68
*Veillonella parvula**	69.7	96.3	7.261	10.72	86.41
*Streptococcus mitis**	60.5	54.2	5.539	8.179	94.58
*Rothia mucilaginosa**	31.6	41.4	3.667	5.415	100

C

Taxon	% Mean abundance		Average dissimilarity	Contribution %	Cumulative %
	Female	Male			
Staphylococcus aureus	120	93.7	10.9	12.02	12.02
Prophorymonas catoniae	74.4	102	9.14	10.09	22.11
***Burkholderia cepacia* complex**	113	68.1	7.622	8.41	30.52
Haemophilus influenzae	53	90.8	7.217	7.963	38.48
Stenotrophomonas maltophilia	64.8	31	4.562	5.033	43.51
Prevotella	39	44.6	4.319	4.765	48.28
Prevotella oris	39.8	19.9	3.26	3.597	51.88
Neissaria flavescens	20.3	28.1	2.539	2.801	54.68
Granulicatella adiacens	7.82	24.9	2.188	2.414	57.09
Prevotella denticola	20.8	16.8	2.146	2.368	59.46
Streptococcus parasanguinis	14.5	17.9	2.067	2.281	61.74
Fusobacterium nucleatum	6.92	18.6	1.976	2.18	63.92
Porphyromonas catoniae	9.75	17.7	1.633	1.802	65.72
Porphyromonas endodontalis	1.98	19.4	1.418	1.565	67.29
Leptotrichia wadei	15.9	11	1.405	1.55	68.84
Actinomyces meyeri	7.56	13.8	1.394	1.538	70.37
Prevotella pallens	12.5	8.22	1.35	1.49	71.86
Megasphaera micronuciformis	9.13	9.82	1.214	1.339	73.2
Haemophilus parainfluenzae	8.9	10.4	1.156	1.275	74.48
Prevotella nanceiensis	5.25	10.6	0.9745	1.075	75.55
Streptococcus infantis	10.2	6.46	0.9417	1.039	76.59

6.4 Discussion

Culture independent technologies have provided a wealth of previously unconsidered information on the bacterial community within the CF lung. However, the power of these studies has been limited by the number of samples investigated. To date studies have relied on small sample sets of less than 50 samples, typically collected from a single centre (3, 6, 8, 9). In their paper published in 2012, Delhaes *et al*, acknowledge that the cohort size they use is small, 45 CF individuals, making it difficult to understand the underpinning relationships between the bacterial community and disease state (6).

Despite the limitations of previous studies, researchers have recognised the importance of understanding how the bacterial community relates to clinical factors and patient well being. Studies by Klepac-Ceraj *et al* (2010), Cox *et al* (2010), van der Gast *et al* (2011) and Delhaes *et al* (2012), investigated the relationship between bacterial diversity, age and lung function (3, 6, 8, 9). While Klepac-Ceraj *et al* (2010) found that in children, although diversity is not correlated with lung function, they did find a strong relationship between bacterial diversity and age using throat swabs from children between 2 and 16 years, analysed using Phylochip hybridisation (8). This suggests that as children age, bacterial diversity decreases (8). Similar findings were seen by Cox *et al* (2010) using a cohort of 51, age stratified patients (from 1 to 72 years old) observing a decrease in lung function with age (3). Smaller sample sets of 14 and 4 patients, van der Gast *et al* (2011) and Delhaes *et al* (2012) failed to show a correlation between age and lung function, however they did identify a strong correlation between bacterial community diversity and lung function (6, 9).

The relationship between age, lung function and bacterial diversity could potentially be an important marker of disease progression in CF patients. However, a high degree of variation was observed in the previously published studies. By investigating this relationship with the much larger cohort of patients involved in this study a more complete understanding of these relationships could be gained. It followed that using the three recognised measures of bacterial diversity; species richness, the Shannon-Wiener index and the inverse Simpson's index, no significant pattern in diversity was observed with age.

Although, no significant trend in diversity was observed with age (mean age 29.9± SD 9.74, max 71, min 14.24), a significant positive relationship was found between diversity and lung function. The failure to identify a significant pattern between diversity and age, in this adult CF cohort is perhaps surprising as longitudinal studies have previously demonstrated that lung function decreases over the life of CF patients (4). However the rate of this decline has been demonstrated to be dependent on the nature of the CF symptoms experienced by an individual patient; with some individuals experiencing a more progressive disease phenotype than others (4). This interpatient variation results in age being a poor predictor of bacterial diversity and lung function; while older patients are in theory likely to have worse lung function and therefore a less diverse microbial community than younger individuals, a patient who survives beyond the average age of mortality (~37 years of age) is likely to have a milder disease phenotype than other individuals. As such, while lung function may be a better predictor of bacterial diversity than age, a more longitudinal approach may be required for individual patients in order to be confident in the results.

Despite the failure to observe any pattern in diversity with age, the relationship between bacterial diversity and lung function has been observed in previous studies. However, while van der Gast *et al* (2011) reports r^2 values of over 0.4 when describing this relationship (9), the variance observed using a significantly larger data set resulted in r^2 values of 0.22, 0.21, 0.17 (for species richness, the Shannon-Wiener index and the inverse Simpson's index respectively). Despite the fact that lung function (as measure by forced expiratory volume in 1 second (FEV_1)) has been found to be the single best predictor of mortality in CF patients (27), the high degree of variation implies that single point measurements of lung function alone are unable to give a satisfactory indication of species diversity.

To further examine the relationship between, age, lung function and the bacterial community, Mantel and partial Mantel test were utilized. Using the Mantel test it was found that community composition, as measured by Bray-Curtis measure of similarity, showed a significant correlation with both age and lung function for the whole community as well as for the common and rare groups. This result suggests that patients with similar bacterial

communities were more likely to be similar in age or lung function. To investigate if these two variables were correlated partial Mantel tests were used, revealing that while controlling for age, FEV_1 was correlated with lung function, this was not the case when FEV_1 was used to control for age. Importantly no autocorrelation was observed between age and lung function (r=0, *P*=0.965). This suggests that similarities in lung function are related to bacterial composition. The observed correlations between lung function and community composition adds weight to the results of the diversity data, and suggests that lung function is strongly associated to the bacterial community and that this is irrespective of age.

Despite relationships observed between the bacterial community and lung function, the variance explained by FEV_1 was still found to be low, suggesting that other clinical factors were involved. In order to investigate additional clinical factors influencing the correlation observed between FEV_1 and community composition partial mantel tests were used. These tests revealed that FEV_1 and clinical status (stable or being treated for CFPE), BMI and liver disease were correlated with similarity in bacterial community composition. Correlation between FEV_1 and clinical status was not surprising as it has previously been shown that CFPE have negative effects on lung function (28). This relationship between the bacterial community and CFPE is further explored in Chapter 7.

The relationship between the bacterial community and clinical status has been explored in several publications (5, 29, 30) however, it is unclear how liver disease is related to the bacterial community within the lung. Investigation into changes in bacterial diversity as a result of this showed no significant relationship, suggesting that although liver disease did not correlate with bacterial diversity it did with community composition. Liver disease may therefore be an area which requires further examination.

Patients with similar heights and weights would be expected to have similar lung capacities. Further, it has also been shown in previous studies that increases in BMI are associated with increases in FEV_1 (31). It therefore makes clinical sense for BMI and FEV_1 to be correlated with community composition. Investigation of the relationship between bacterial diversity and BMI revealed that BMI had a significant positive relationship between the common group

bacterial diversity. This relationship however, explained very little of the variance (*S**; r^2=0.039, *H'*; $r^2$0.028 ,1-*D*; r^2=0.028), suggesting that while changes in BMI are associated with changes in the bacterial community other factors are effecting this relationship, one of these is likely to be FEV_1.

The high levels of interpatient variation observed when considering the bacterial community in relation to age and lung function, suggests that other factors may be playing a significant role in patient outcomes. Previous studies have implied that CFTR genotype may have potentially important consequences for disease progression (10). While the majority of patients were found to ΔF508 homozygous, the huge variation in other mutations seen in the non-homozygous individuals resulted in it being impossible to look for trends due to specific mutations. As a result, and following the convention outlined by Klepac-Ceraj *et al* (2010), data was partitioned into 3 groups; ΔF508 homozygotes, ΔF508 heterozygotes and other mutations. Whilst results of the investigation by Klepac-Ceraj *et al* (2010) suggested a significant difference in the bacterial community between patients with ΔF508 mutations and other mutations (8), our results show no significant change in bacterial diversity between genotype groups. Although ANOSIM, using Bray-Curtis similarity index revealed a significant difference (*P*=0.03) in community composition between patients with homozygous and heterozygous ΔF508 mutations. This was only observed when the whole community was investigated. Mantels tests on the other hand, revealed no difference in bacterial composition. While there is still a probability that genotype plays a role in disease progression, our results suggest that it does not have a significant effect on bacterial diversity, although some evidence was observed that there might be some changes in community composition. In order to advance our understanding of the impact genotype plays on the bacterial community a much larger cohort of rare genotypes would be required.

Historically the survival rates of CF patients have shown that female patients are likely to die before male (11, 12). In recent years this gap has been shown to close (13) however, there still remains the question of what causes this disparity between genders. In order to investigate if the bacterial community differs between male and female patients samples

were partitioned by gender and bacterial diversity and composition. The comparison of bacterial diversity using ANOVA revealed only a significant difference in the inverse Simpson's index, indicating that female patients have a lower diversity when accounting for species presence and abundance. Even so, this finding accounted for very little of the variance observed (r^2=0.02).

To further investigate any relationship between the bacterial community and gender, analysis of similarity was employed. ANOSIM of the whole community revealed significant differences in bacterial similarity as calculated by Sørensens similarity index however, this was not observed when Bray-Curtis similarity index was used. This result suggests that while the members of the bacterial community may be similar, bacterial abundances vary between genders. To investigate if the common or rare taxa groups were responsible for these changes ANOSIMs were carried on the partitioned data. This revealed no significant difference in the common group OTUs between females and males using either Sørensens or Bray-Curtis similarity indices. However importantly, when the rare group was considered both similarity indices revealed significant differences in bacterial community composition between genders.

Similarity percentage analysis allowed the percentage contribution of OTUs to the female (mean age 28.7± SD 9.76, max 67, min 14.24) and male (mean age 30.7± SD 9.68, max 71, min 17.38) communities to be calculated. *Pseudomonas aeruginosa* was shown to contribute most to bacterial similarity; it was however shown to have a higher mean abundance in females than males. Interestingly *Staphylococcus aureus*, which was found to be the third biggest contributor to the whole community similarity (6.7%) and largest contributor to the rare group similarity (12.02%), were also found to have a larger mean abundance in females than males, suggesting that these recognised CF pathogens are more abundant in female CF patients. This relationship requires further investigation as it has been shown that co-infection of *P. aeruginosa* and *S. aureus* may be related to worse disease outcomes (32). The relationship between gender and the bacterial community may be an important avenue for further study as it may signify the requirement for gender tailored antibiotic intervention.

A published study by Stressmann *et al* (2011) comparing the bacterial community between 2 distinct CF centres, showed that while the common bacterial community was found to be similar between sites, samples collected from the US centre was significantly less diverse (15). The large number of samples and centres involved in this project allowed the investigation of between centres differences. By investigating diversity and composition at the metacommunity level across all study centres, this work could identify the effect of centre and geographic location on the bacterial community within the lung. Investigation of the whole community diversity alongside that of the common and rare groups revealed a significant difference in bacterial diversity in the whole and common groups, although the rare community showed only a significant change in the values of richness. However, while overall changes in diversity by centre are shown to be significantly different, post-hoc testing found few significant differences between centres. This suggests that while centre may be a contributing factor to bacterial diversity, differences in interpatient variation is likely to be more to account for more of the observed changes.

Mantel tests showed community composition, as measured by Bray-Curtis similarity index, was significantly correlated with distance between centres. Showing that centres closer together geographically showed more similarity in community composition. To further investigate these changes ANOSIMs were carried out, using both Bray-Curtis and Sørensen indices of similarity, revealing significant changes in bacterial composition between centres. Despite the significant result of the Mantel tests, ANOSIMs on the whole community as well as the common and rare groups revealed significant differences between centres. In most cases this was irrespective of geographical distance. The common OTU group was found to be more conserved between centres than the rare group, suggesting that differences due to, centre, geographical distance or ethnic origin, were driven by the rare community. These results indicate that while geographical difference is a contributing factor to differences in the bacterial community, centre related differences such as treatment regimes also play a significant role in bacterial community composition.

This study has revealed a range of clinical factors associated with differences in the bacterial community of the CF lung. The basic statistical analyses carried out here have provided an overview of the potentially important relationships which warrant further investigation. Previous studies have suggested that relationships between clinical factors such as, age, FEV1, CFTR genotype and gender could have potential clinical implications however, the sample sizes involved in these studies have failed to observe the true level of interpatient variation.

This study is an initial investigation into a large and complex dataset, the size of which has never before been used in the exploration of the microbial community within the CF lung. As such, this data has the potential to be further mined as important clinical relationships are revealed. This study aimed to validate previously observed relationships between clinical factors and the bacterial community within the CF lung by testing them against this unparalleled dataset. While the size of this dataset provides unrivalled statistical power, it also adds levels of complexity to the analysis. Results from this study have so far revealed that no single clinical factor can explain the variance observed between patients.

The size of this dataset, and the associated metadata collected alongside, provides the potential for a wide range of further investigations. Some of the future work on this data will include the use of more in depth and multivariate analyses in an attempt to explain the bacterial community within the lung and how this affects disease progression. On top of this, the role antibiotic regimes play in the bacterial community observed will be investigated to identify if particular antibiotic combinations result in better patient outcomes. In addition, further investigation into the role gender plays in bacterial community composition will be carried out. To gain a greater understanding of poor disease outcomes in female CF patients their bacterial community needs to be investigated further, particularly the presence of *Pseudomonas aeruginosa* and *Staphylococcus aureus* in the lower airways. This will be discussed more fully in Chapter 8.

6.5 References

1. **Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ, Bruce KD.** 2003. Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. J Clin Microbiol **41:**3548-3558.

2. **Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, Connett GJ, Bruce KD.** 2006. Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. J Clin Microbiol **44:**2601-2604.

3. **Cox M, Allgaier M, Taylor B, Baek M, Huang Y, Daly R, Karaoz U, Andersen G, Brown R, Fujimura K, Wu B, Tran D, Koff J, Kleinhenz M, Nielson D, Brodie E, Lynch S.** 2010. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. PLOS One **5:**e11044.

4. **Zhao JC, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ.** 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. PNAS USA **109:**5809-5814.

5. **Fodor A, Klem E, Gilpin D, Elborn J, Boucher R, Tunney M, Wolfgang M.** 2012. The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. PLOS One **7:**e45001.

6. **Delhaes L, Monchy S, Frealle E, Hubans C, Salleron J, Leroy S, Prevotat A, Wallet F, Wallaert B, Dei-Cas E, Sime-Ngando T, Chabe M, Viscogliosi E.** 2012. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. PLOS One **7:**e36313.

7. **Dodge JA, Lewis PA, Stanton M, Wilsher J.** 2007. Cystic fibrosis mortality and survival in the UK: 1947-2003. Eur Respir J **29:**522-526.

8. **Klepac-Ceraj V, Lemon KP, Martin TR, Allgaier M, Kembel SW, Knapp AA, Lory S, Brodie EL, Lynch SV, Bohannan BJM, Green JL, Maurer BA, Kolter R.** 2010. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. Environ Microbiol **12:**1293-1303.

9. **van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD.** 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. ISME J **5:**780-791.

10. **Zielenski J, Tsui LC.** 1995. Cystic fibrosis: Genotypic and phenotypic variations. Annu Rev Genet **29:**777-807.

11. **McKone EF, Goss CH, Aitken ML.** 2006. CFtr genotype as a predictor of prognosis in cystic fibrosis*. Chest **130:**1441-1447.

12. **Rosenfeld M, Davis R, FitzSimmons S, Pepe M, Ramsey B.** 1997. Gender gap in cystic fibrosis mortality. Am J Epidemiol **145:**794-803.

13. **Verma N, Bush A, Buchdahl R.** 2005. Is there still a gender gap in cystic fibrosis? Chest **128:**2824-2834.

14. **Johansen HK, Kovesi TA, Koch C, Corey M, Hoiby N, Levison H.** 1998. Pseudomonas aeruginosa and Burkholderia cepacia infection in cystic fibrosis patients treated in Toronto and Copenhagen. Pediatr Pulmonol **26:**89-96.

15. **Stressmann FA, Rogers GB, Klem ER, Lilley AK, Donaldson SH, Daniels TW, Carroll MP, Patel N, Forbes B, Boucher RC, Wolfgang MC, Bruce KD.** 2011. Analysis of the bacterial communities present in lungs of patients with cystic fibrosis from American and British centers. J Clin Microbiol **49:**281-291.

16. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics (Oxford, England) **27:**2194-2200.

17. **Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR.** 1996. The Ribosomal Database Project (RDP). Nucleic Acids Res **24:**82-85.

18. **Schloss PD, Handelsman J.** 2006. Toward a census of bacteria in soil. PLOS Comput Biol **2:**786-793.

19. **Salter S, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman N, Walker AW.** 2014. Reagent contamination can critically impact sequence-based microbiome analyses. bioRxiv **doi:** http://dx.doi.org/10.1101/007187

20. **The R Core Development Team.** 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

21. **Magurran AE, Henderson PA.** 2003. Explaining the excess of rare species in natural species abundance distributions. Nature **422:**714-716.

22. **Oksanen J, F Guillaume Blanchet, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H.** 2013. vegan: Community Ecology Package. R package version 2.0-7.

23. **Hammer Ø, Harper, D.A.T., Ryan, P.D.** 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron. **4:**9pp.

24. **Clarke KR.** 1993. Nonparametric Multivariate Analyses of Changes in Community Structure. Aust J Ecol **18:**117-143.

25. **Mantel N.** 1967. The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res **27:**209-220.

26. **Hazard C, Gosling P, van der Gast CJ, Mitchell DT, Doohan FM, Bending GD.** 2013. The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. ISME J **7:**498-508.

27. **Sanders DB, Bittner RCL, Rosenfeld M, Hoffman LR, Redding GJ, Goss CH.** 2010. Failure to Recover to Baseline Pulmonary Function after Cystic Fibrosis Pulmonary Exacerbation. Am J Resp Crit Care **182:**627-632.

28. **Daniels T, Rogers G, Stressmann F, van der Gast C, Bruce K, Jones G, Connett G, Legg J, Carroll M.** 2013. Impact of antibiotic treatment for pulmonary exacerbations on bacterial diversity in cystic fibrosis. J Cyst Fibros **12:**22 - 28.

29. **Carmody LA, Zhao J, Schloss PD, Petrosino JF, Murray S, Young VB, Li JZ, LiPuma JJ.** 2013. Changes in Cystic Fibrosis Airway Microbiota at Pulmonary Exacerbation. Ann Am Thorac Soc **10:**179-187.

30. **Stephenson AL, Mannik LA, Walsh S, Brotherwood M, Robert R, Darling PB, Nisenbaum R, Moerman J, Stanojevic S.** 2013. Longitudinal trends in nutritional status and the relation between lung function and BMI in cystic fibrosis: a population-based cohort study. Am J Clin Nutr **97:**872-877.

31. **Korgaonkar A, Trivedi U, Rumbaugh KP, Whiteley M.** 2013. Community surveillance enhances Pseudomonas aeruginosa virulence during polymicrobial infection. PNAS **110:**1059-1064.

32. **Sibley CD, Duan K, Fischer C, Parkins MD, Storey DG, Rabin HR, Surette MG.** 2008. Discerning the Complexity of Community Interactions Using a *Drosophila* Model of Polymicrobial Infections. PLOS Pathog **4:**e1000184.

6.6 Appendix

Table A6.1 Species-level identities of detected bacterial taxa identified from 311 sputum samples.

Sputum collected from CF patients (*n*=292) and Non-CF controls (*n*=19). Given the length of the ribosomal sequences analysed (approx 350bp) these identities should be considered putative. Taxa partitioned as common are shown in bold.

Class	Family	Taxon name
Acidobacteria	Acidobacteria Gp1 family incertae sedis	*Granulicella mallensis*
	Acidobacteria Gp16 family incertae sedis	*Acidobacteria*
		Dehalococcoides ethenogenes
		Paenibacillus taichungensis
Actinobacteria	Acidimicrobineae	*Aciditerrimonas ferrireducens*
	Actinomycetaceae	*Actinomyces*
		Actinomyces cardiffensis
		Actinomyces dentalis
		Actinomyces graevenitzii
		Actinomyces hongkongensis
		Actinomyces israelii
		Actinomyces johnsonii
		Actinomyces massiliensis
		Actinomyces meyeri
		Actinomyces naeslundii
		Actinomyces odontolyticus
		Actinomyces oris
		Actinomyces radicidentis
		Actinomyces viscosus
		Arcanobacterium haemolyticum
	Actinomycetales	*Micromonospora*
	Bifidobacteriaceae	*Alloscardovia omnicolens*
		Bifidobacterium animalis
		Bifidobacterium breve
		Bifidobacterium dentium
		Bifidobacterium longum
		Bifidobacterium subtile
		Bombiscardovia coagulans

Table A6.1 Continued

Class	Family	Taxon name
		Parascardovia denticolens
		Scardovia wiggsiae
		Brevibacterium epidermidis
		Brevibacterium massiliense
		Brevibacterium otitidis
		Brevibacterium sanguinis
	Cellulomonadaceae	*Cellulomonas*
		Tropheryma whipplei
	Conexibacteraceae	*Conexibacter woesei*
	Coriobacteriaceae	*Atopobium parvulum*
		Atopobium rimae
		Cryptobacterium curtum
		Olsenella uli
	Corynebacteriaceae	*Corynebacterium afermentans*
		Corynebacterium amycolatum
		Corynebacterium casei
		Corynebacterium diphtheriae
		Corynebacterium durum
		Corynebacterium lipophiloflavum
		Corynebacterium matruchotii
		Corynebacterium mucifaciens
		Corynebacterium pseudodiphtheriticum
		Corynebacterium striatum
		Corynebacterium tuberculostearicum
		Corynebacterium ureicelerivorans
		Corynebacterium variabilis
		Turicella otitidis
	Demequinaceae	*Demequina aestuarii*
	Dermacoccaceae	*Dermacoccus nishinomiyaensis*
		Kytococcus sedentarius
	Dietziaceae	*Dietzia maris*
	Geodermatophilaceae	*Geodermatophilus*
	Intrasporangiaceae	*Janibacter*
		Knoellia melonis
		Ornithinimicrobium murale
		Phycicoccus
	Microbacteriaceae	*Curtobacterium*

Table A6.1 Continued

Class	Family	Taxon name
		Curtobacterium flaccumfaciens
		Leifsonia bigeumensis
		Leucobacter
		Leucobacter albus
		Leucobacter chromiiresistens
		Microbacterium arborescens
		Microbacterium sediminicola
		Plantibacter flavus
	Micrococcaceae	*Arthrobacter*
		Kocuria palustris
		Kocuria rhizophila
		Micrococcus luteus
		Rothia dentocariosa
		Rothia mucilaginosa
	Micromonosporaceae	*Catellatospora bangladeshensis*
	Mycobacteriaceae	*Mycobacterium massiliense*
		Mycobacterium obuense
	Nakamurellaceae	*Humicoccus flavidus*
	Nocardiaceae	*Gordonia terrae*
		Nocardia farcinica
		Rhodococcus
	Nocardioidaceae	*Nocardioides*
	Promicromonosporaceae	*Promicromonospora*
		Xylanibacterium ulmi
	Propionibacteriaceae	*Brooklawnia massiliensis*
		Microlunatus phosphovorus
		Propionibacterium acidifaciens
		Propionibacterium acnes
		Propionibacterium propionicum
		Propionicimonas paludicola
	Sanguibacteraceae	*Sanguibacter inulinus*
	Solirubrobacteraceae	*Patulibacter medicamentivorans*
	Streptomycetaceae	*Streptomyces*
	Thermomonosporaceae	*Thermomonospora*
Alphaproteobacteria	Acetobacteraceae	*Acidomonas methanolica*
		Asaia
		Granulibacter bethesdensis
		Roseomonas mucosa

Table A6.1 Continued

Class	Family	Taxon name
		Roseomonas riguiloci
	Beijerinckiaceae	*Methylocella palustris*
	Brucellaceae	*Bacillus malacitensis*
		Ochrobactrum intermedium
	Caulobacteraceae	*Brevundimonas*
		Caulobacter fusiformis
		Caulobacter vibrioides
	Erythrobacteraceae	*Erythrobacter*
		Porphyrobacter
	Hyphomicrobiaceae	*Devosia riboflavina*
		Rhodomicrobium vannielii
		Shinella
	Methylobacteriaceae	*Methylobacterium*
	Phyllobacteriaceae	*Shinella zoogloeoides*
	Rhodobacteraceae	*Leisingera aquaemixtae*
		Paracoccus alcaliphilus
		Paracoccus aminophilus
		Paracoccus aminovorans
		Paracoccus denitrificans
		Paracoccus homiensis
		Paracoccus yeei
		Rubellimicrobium mesophilum
		Rubellimicrobium roseum
	Rhodospirillaceae	*Azospirillum amazonense*
		Inquilinus limosus
	Rhodospirillales	*Reyranella massiliensis*
	Sphingomonadaceae	*Novosphingobium acidiphilum*
		Novosphingobium pentaromativorans
		Sphingomonas
	Xanthobacteraceae	*Orientia tsutsugamushi*
Bacilli	Aerococcaceae	*Abiotrophia defectiva*
	Bacillaceae 1	*Bacillus circulans*
		Bacillus longiquaesitum
		Bacillus pumilus
	Bacillales Incertae Sedis XI	*Gemella bergeri*
		Gemella haemolysans
		Gemella morbillorum

Table A6.1 Continued

Class	Family	Taxon name
		Gemella sanguinis
		Exiguobacterium
	Carnobacteriaceae	*Dolosigranulum pigrum*
		Granulicatella adiacens
		Granulicatella elegans
	Enterococcaceae	*Enterococcus*
		Enterococcus faecalis
		Enterococcus faecium
		Enterococcus malodoratus
		Enterococcus moraviensis
		Enterococcus saccharolyticus
	Lactobacillaceae	*Lactobacillus acidophilus*
		Lactobacillus amylolyticus
		Lactobacillus casei
		Lactobacillus delbrueckii
		Lactobacillus fermentum
		Lactobacillus gasseri
		Lactobacillus helveticus
		Lactobacillus iwatensis
		Lactobacillus kimbladii
		Lactobacillus melliventris
		Lactobacillus mucosae
		Lactobacillus mucosae
		Lactobacillus nantensis
		Lactobacillus nasuensis
		Lactobacillus oris
		Lactobacillus plantarum
		Lactobacillus rhamnosus
		Lactobacillus salivarius
		Lactobacillus vaginalis
		Pediococcus acidilactici
	Leuconostocaceae	*Fructobacillus*
	Listeriaceae	*Brochothrix thermosphacta*
	Planococcaceae	*Lysinibacillus fusiformis*
	Staphylococcaceae	*Staphylococcus aureus*
		Staphylococcus auricularis
		Staphylococcus epidermidis
		Staphylococcus pasteuri

Table A6.1 Continued

Class	Family	Taxon name
	Streptococcaceae	*Lactococcus lactis*
		Streptococcus
		Streptococcus agalactiae
		Streptococcus anginosus
		Streptococcus australis
		Streptococcus constellatus
		Streptococcus cristatus
		Streptococcus dysgalactiae
		Streptococcus gordonii
		Streptococcus infantis
		Streptococcus intermedius
		Streptococcus mitis
		Streptococcus mutans
		Streptococcus oralis
		Streptococcus parasanguinis
		Streptococcus pyogenes
		Streptococcus salivarius
		Streptococcus sobrinus
Bacteroidetes	Flavobacteriales	*Sediminibacter*
Bacteroidia	Bacteroidaceae	*Bacteroides*
		Bacteroides acidifaciens
	Bacteroidalesincertaesedis	*Phocaeicola abscessus*
	Porphyromonadaceae	*Butyricimonas faecihominis*
		Coprobacter fastidiosus
		Dysgonomonas mossii
		Odoribacter splanchnicus
		Parabacteroides goldsteinii
		Porphyromonas asaccharolytica
		Porphyromonas bennonis
		Porphyromonas catoniae
		Porphyromonas endodontalis
		Porphyromonas gingivalis
		Porphyromonas macacae
		Porphyromonas somerae
		Prophorymonas catoniae
		Tannerella forsythia
	Prevotellaceae	*Alloprevotella rava*
		Gordonia jinhuaensis

Table A6.1 Continued

Class	Family	Taxon name
		Prevotella
		Prevotella bivia
		Prevotella buccae
		Prevotella dentalis
		Prevotella denticola
		Prevotella enoeca
		Prevotella fusca
		Prevotella histicola
		Prevotella intermedia
		Prevotella loescheii
		Prevotella melaninogenica
		Prevotella micans
		Prevotella multiformis
		Prevotella nanceiensis
		Prevotella nigrescens
		Prevotella oralis
		Prevotella oris
		Prevotella oulorum
		Prevotella pallens
		Prevotella saccharolytica
		Prevotella salivae
		Prevotella scopos
		Prevotella shahii
		Prevotella tannerae
		Prevotella veroralis
	Rikenellaceae	*Alistipes shahii*
		Rikenella microfusus
Betaproteobacteria	Alcaligenaceae	*Achromobacter xylosoxidans*
		Alcaligenes faecalis
		Bordetella petrii
		Parapusillimonas granuli
		Pusillimonas
	Burkholderiaceae	*Burkholderia cepacia* complex
		Cupriavidus respiraculi
		Ralstonia
		Ralstonia mannitolilytica
		Ralstonia pickettii
	Comamonadaceae	*Acidovorax wautersii*

Table A6.1 Continued

Class	Family	Taxon name
		Lautropia mirabilis
		Ottowia
		Ramlibacter tataouinensis
		Variovorax paradoxus
		Methylotenera versatilis
	Neisseriaceae	*Eikenella corrodens*
		Kingella denitrificans
		Kingella oralis
		Neissaria flavescens
		Neisseria bacilliformis
		Neisseria elongata
		Neisseria flavescens
		Neisseria lactamica
		Neisseria mucosa
		Neisseria oralis
		Neisseria pharyngis
		Neisseria sicca
	Oxalobacteraceae	*Herbaspirillum huttiense*
		Massilia aurea
		Massilia timonae
	Rhodocyclaceae	*Dechloromonas agitata*
		Methyloversatilis universalis
		Propionivibrio
	Sutterellaceae	*Sutterella parvirubra*
Clostridia	Clostridiaceae 1	*Clostridium paraputrificum*
	Clostridiales	*Clostridiales*
	Clostridiales Family XI. Incerta Sedis	*Peptoniphilus*
	Clostridiales Family XIII. Incertae Sedis	*Eubacterium sulci*
		Mogibacterium timidum
	Clostridiales Incertae Sedis XI	*Anaerococcus*
		Parvimonas micra
		Peptoniphilus lacrimalis
		Peptoniphilus tyrrelliae
	Clostridiales Incertae Sedis XIII	*Mogibacterium diversum*
	Eubacteriaceae	*Eubacterium brachy*
		Eubacterium infirmum

Table A6.1 Continued

Class	Family	Taxon name
		Eubacterium nodatum
		Eubacterium saphenum
		Eubacterium uniforme
	Incertae Sedis XI	*Peptoniphilus asaccharolyticus*
	Lachnospiraceae	*Butyrivibrio fibrisolvens*
		Butyrivibrio hungatei
		Catonella morbi
		Clostridium algidixylanolyticum
		Clostridium clostridioforme
		Clostridium hathewayi
		Clostridium symbiosum
		Clostridium xylanovorans
		Eubacterium oxidoreducens
		Howardella ureilytica
		Johnsonella ignava
		Lachnospiraceae
		Moryella indoligenes
		Oribacterium sinus
		Shuttleworthia satelles
	Peptococcaceae	*Peptococcus*
	Peptostreptococcaceae	*Clostridium metallolevans*
		Clostridium rectum
		Eubacterium yurii
		Filifactor alocis
	Ruminococcaceae	*Clostridium methylpentosum*
		Clostridium orbiscindens
		Faecalibacterium prausnitzii
		Oscillibacter ruminantium
		Ruminococcus
Cyanobacteria	Chroococcales	*Synechococcus elongatus*
Deferribacteres	Deferribacteraceae	*Mucispirillum schaedleri*
Deltaproteobacteria	Cystobacteraceae	*Cystobacter miniatus*
	Desulfobulbaceae	*Desulfobulbus*
	Desulfomicrobiaceae	*Desulfomicrobium orale*
	Polyangiaceae	*Chondromyces apiculatus*
		Sorangium cellulosum
Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*
		Campylobacter gracilis

Table A6.1 Continued

Class	Family	Taxon name
		Campylobacter mucosalis
		Campylobacter rectus
		Campylobacter showae
		Campylobacter sputorum
	Helicobacteraceae	*Sulfuricurvum kujiense*
Erysipelotrichia	Erysipelotrichaceae	*Bulleidia extructa*
		Coprobacillus catenaformis
		Erysipelothrix tonsillarum
		Lactobacillus catenaformis
		Solobacterium moorei
Flavobacteria	Cryomorphaceae	*Salinirepens amamiensis*
	Flavobacteriaceae	*Bergeyella zoohelcum*
		Capnocytophaga gingivalis
		Capnocytophaga granulosa
		Capnocytophaga haemolytica
		Capnocytophaga leadbetteri
		Capnocytophaga ochracea
		Capnocytophaga sputigena
		Chryseobacterium anthropi
		Elizabethkingia miricola
		Flavobacterium
		Kocuria rosea
		Myroides odoratus
		Wautersiella falsenii
		Myroides odoratimimus
		Myroides profundi
Fusobacteria	Fusobacteriaceae	*Fusobacterium necrophorum*
		Fusobacterium nucleatum
		Fusobacterium periodonticum
	Leptotrichiaceae	*Leptotrichia buccalis*
		Leptotrichia goodfellowii
		Leptotrichia hofstadii
		Leptotrichia hongkongensis
		Leptotrichia shahii
		Leptotrichia trevisanii
		Leptotrichia wadei
		Sneathia sanguinegens
		Streptobacillus moniliformis

Table A6.1 Continued

Class	Family	Taxon name
Gammaproteobacteria	Aeromonadaceae	*Aeromonas*
	Cardiobacteriaceae	*Cardiobacterium hominis*
	Coxiellaceae	*Legionella spiritensis*
	Enterobacteriaceae	*Arsenophonus nasoniae*
		Erwinia rhapontici
		Escherichia coli
		Klebsiella oxytoca
		Morganella morganii
		Pantoea ananatis
		Proteus mirablis
		Providencia alcalifaciens
		Providencia rettgeri
		Serratia marcescens
	Gammaproteobacteria family incertae sedis	*Orbus hercynius*
		Orbus sasakiae
		Psychromonas marina
	Halomonadaceae	*Halomonas*
	Legionellaceae	*Legionella fairfieldensis*
	Moraxellaceae	*Alkanindiges illinoisensis*
		Enhydrobacter aerosaccus
		Moraxella catarrhalis
		Moraxella macacae
		Moraxella nonliquefaciens
	Pasteurellaceae	*Aggregatibacter actinomycetemcomitan*
		Aggregatibacter segnis
		Haemophilus haemolyticus
		Haemophilus influenzae
		Haemophilus parahaemolyticus
		Haemophilus parainfluenzae
		Haemophilus sputorum
	Pseudomonadaceae	***Pseudomonas aeruginosa***
	Sinobacteraceae	*Povalibacter uvarum*
	Xanthomonadaceae	*Ignatzschineria ureiclastica*
		Stenotrophomonas maltophilia
		Thermomonas hydrothermalis
		Xanthomonas
Gemmatimonadetes	Gemmatimonadaceae	*Gemmatimonas aurantiaca*

Table A6.1 Continued

Class	Family	Taxon name
Ignavibacteria	Ignavibacteriaceae	*Ignavibacterium album*
Mollicutes	Mycoplasmataceae	*Mycoplasma faucium*
		Mycoplasma feliminutum
		Mycoplasma salivarium
Negativicutes	Veillonellaceae	*Anaeroglobus geminatus*
		Centipeda periodontii
		Dialister invisus
		Dialister micraerophilus
		Dialister pneumosintes
		Megasphaera micronuciformis
		Mitsuokella
		Schwartzia succinivorans
		Selenomonas
		Selenomonas artemidis
		Selenomonas dianae
		Selenomonas flueggei
		Selenomonas infelix
		Selenomonas noxia
		Selenomonas sputigena
		Veillonella atypica
		Veillonella dispar
		Veillonella parvula
Planctomycetacia	Planctomycetaceae	*Gemmata obscuriglobus*
		Telmatocola sphagniphila
Spartobacteria	Spartobacteria family incertae sedis	*Chthoniobacter flavus*
Sphingobacteria	Chitinophagaceae	*Chitinophaga*
		Chitinophaga niastensis
		Sediminibacterium
	Cytophagaceae	*Hymenobacter*
	Sphingobacteriaceae	*Pedobacter cryoconitis*
		Pedobacter heparinus
		Sphingobacterium
Spirochaetes	Spirochaetaceae	*Treponema*
		Treponema amylovorum
		Treponema denticola
		Treponema maltophilum
		Treponema putidum

Table A6.1 Continued

Class	Family	Taxon name
		Treponema socranskii
		Treponema vincentii
SR1 class incertae sedis	SR1 family incertae sedis	SR1 bacterium human oral taxon
Synergistetes	Synergistetes	*Synergistetes*
TM7 class incertae sedis	TM7 family incertae sedis	TM7 phylum sp. oral taxon

Table A6.2 Mantel test analyses for the investigation of autocorrelation.

The *P*-values have been calculated using the distribution of the Mantel test statistic (*r*) estimated for 9999 permutations. *N*=152, *n*= 11476 pair wise comparisons between groups. Significant *P*-values are indicated in bold. A, Mantel tests carried out using a two tailed test, B, Mantel tests carried out using a lower tailed test

A

Matrix A	Matrix B	r	*P*
Age	BMI	-0.045	**<0.001**
Age	Clinical status	-0.037	**<0.001**
Age	Diabetes	0.044	**<0.001**
Age	Pancreatic	0.013	0.168
Age	Gender	0.005	0.559
Age	liver disease	0.012	0.212
BMI	Clinical status	-0.021	**0.027**
BMI	diabetes	-0.002	0.829
BMI	Liver disease	0.097	**<0.001**
BMI	Pancreatic	0.001	0.947
Clinical status	Diabetes	-0.001	0.897
Clinical status	Liver disease	-0.006	0.564
Clinical status	Pancreatic	0.039	**<0.001**
Diabetes	Liver disease	0.025	**0.008**
Diabetes	Pancreatic	0.063	**<0.001**
FEV_1	Age	0	0.965
FEV_1	Gender	0.017	0.071
FEV_1	BMI	0.026	**0.005**
FEV_1	Clinical status	0.014	0.142
FEV_1	Diabetes	0.029	**0.002**
FEV_1	Pancreatic	0.011	0.23
FEV_1	Liver disease	0.062	**<0.001**
Liver disease	Pancreatic	-0.004	0.644
Location	BMI	-0.005	0.616
Location	Clinical status	0.009	0.352
Location	Diabetes	0.107	**<0.001**
Location	FEV_1	-0.012	0.216
Location	Liver disease	-0.002	0.878
Location	Pancreatic	0.255	**<0.001**
Location	Gender	0.008	0.401

Table A6.2 Continued

A

Matrix A	Matrix B	r	*P*
Gender	BMI	0.02	**0.038**
Gender	Diabetes	0	0.99
Gender	Liver disease	-0.011	0.237
Gender	Pancreatic	-0.006	0.541
Gender	Clinical status	0	0.99

B

Matrix A	Matrix B	r	*P*
FEV_1	Liver disease	0.062	1
FEV_1	BMI	0.026	0.997

Table A6.3 Comparison of community composition using analysis of similarity (ANOSIM) for each CFTR genotype.

ANOSIM was carried out using both, A, Sørensen and, B, Bray-Curtis measures of similarity, for the whole, common and rare communities. Matrices below show *R* values in the lower half and *P* values on the upper. Significant *R* values indicate the samples within a group are more similar than would be expected by random chance, therefore the two groups can be considered significantly different. *P* values less than 0.05 were considered significant. Significant *P* values are highlight in green.

A

Whole	Homozygous	Heterozygous	Other
Homozygous	r / *P*	0.5548	0.2777
Heterozygous	-0.002862	r / *P*	0.281
Other	0.0397	0.04064	r / *P*
Common	Homozygous	Heterozygous	Other
Homozygous	r / *P*	0.3812	0.5845
Heterozygous	0.003042	r / *P*	0.6543
Other	-0.02109	-0.03657	r / *P*
Rare	Homozygous	Heterozygous	Other
Homozygous	r / *P*	0.5336	0.3389
Heterozygous	-0.002539	r / *P*	0.3181
Other	0.0263	0.03184	r / *P*

B

Whole	Homozygous	Heterozygous	Other
Homozygous	r / *P*	0.0379	0.9763
Heterozygous	0.02338	r / *P*	0.9972
Other	-0.1143	-0.1527	r / *P*
Common	Homozygous	Heterozygous	Other
Homozygous	r / *P*	0.1385	0.9709
Heterozygous	0.01179	r / *P*	0.9919
Other	-0.1047	-0.1131	r / *P*
Rare	Homozygous	Heterozygous	Other
Homozygous	r / *P*	0.2566	0.5269
Heterozygous	0.007679	r / *P*	0.6378
Other	-0.007451	-0.02893	r / *P*

Chapter 7: Bacterial community dynamics of cystic fibrosis lung infections through changing disease states

Bacterial community dynamics of cystic fibrosis lung infections through changing disease states

7.1 Introduction

It is well established that chronic lung infections are the main cause of morbidity and mortality in cystic fibrosis (CF) patients (1). In recent years, increased use of culture independent molecular technologies has enhanced our understanding of the complex and diverse polymicrobial community associated with CF lung infections (2-5). This improved understanding has lead to better patient management (6), however, there are still gaps in our understanding of how the microbial community relates to disease progression.

Pulmonary symptoms in CF begin in early life, over time a combination of impaired mucociliary clearance, innate immune responses and microbial infection lead to a progressive loss of lung function (7). This gradual decline in pulmonary function is interspersed with periods of acute worsening of respiratory symptoms known as CF pulmonary exacerbations (CFPE) (8). These periods of CFPE are associated with more rapid disease progression and reduced survival, as well as a recognised reduction in quality of life and an increase in overall healthcare costs (9).

Despite efforts, no strict guidelines are in place to describe when patients are experiencing a CFPE. This lack of consensus is mainly due to inconsistency in the symptoms experienced by CF individuals of different ages and disease state, thus making it challenging for paediatricians and adult physicians to come to an agreement (10). Although no consensus for a generally applicable definition has been reached, clinical features include; decreased exercise tolerance, increased cough, increased sputum production, shortness of breath, chest pain, absence from school or work, increased adventitial sounds on lung examination, decreased appetite or weight loss and a decline in lung function. Surprisingly a study by Rosenfeld *et al* (2001) revealed that a change in lung function showed little effect on the sensitivity or specificity of CFPE diagnosis, however, it is still widely relied upon as an indicator of worsening symptoms (11).

The dynamic nature of CF lung disease makes the identification of CFPE particularly challenging, ultimately it is the job of the treating clinician to determine if the patient is experiencing a CFPE and therefore requires intervention. CFPE are treated with aggressive antibiotic therapy, which in the majority of cases involves high dose intravenous (IV) antibiotics targeted at the most abundant organism within the lung; in many cases this is *Pseudomonas aeruginosa* (10).

Efficient intervention is important, as it has been well established that CFPE have important negative effects on pulmonary function, with lung function tests failing to recover to baseline after treatment in one in four cases (12). However, Sanders *et al* (2012) found that quick efficient management of CFPE results in improved prognosis and increased likelihood of lung function recovery, thus underlining the need to efficiently recognise CFPE (12).

To date the majority of studies carried out have involved cross-sectional sampling, where a single sample is taken from each patient within the study (2, 13, 14). These studies have been invaluable for investigating the microbial community within the lung, revealing high levels of microbial diversity and increasing our understanding of the correlations between the lung community and disease progression. Although extremely useful, these studies reveal little of the how community dynamics over time relates to clinical outcomes. For this reason more studies are being carried out using longitudinal sampling with multiple samples from each patient (5, 15-19). By examining the microbial community over time, insights into how the community relates to factors that result in poor clinical outcomes, such as pulmonary exacerbation (18) and its treatment (15) can be investigated.

Longitudinal studies provide useful information about how community changes relate to disease state and provide information on potential targets and biomarkers that could lead to more effective treatment options. In a 10 year study, Zhao *et al* (2011) using 16S rRNA gene pyrosequencing, revealed that bacterial community diversity within the CF lung can be maintained in over prolonged periods in patients with mild disease symptoms. However, in those with more progressive lung disease, diversity is found to decrease significantly over

time (5). Over the shorter scale of one year, Stressmann *et al* (2012) used terminal restriction fragment length polymorphism (T-RFLP) to reveal that the bacterial community was conserved over time and resistant to perturbations (16).

These long term studies have provided valuable information about how changes in the bacterial community composition relate to disease state over time. However they give no indication of how perturbations within the lungs, such as CFPE, relate to patient well being in the short term. Studies by Fodor *et al* (2012) and Carmody *et al* (2013) used high throughput sequencing to investigate changes in bacterial community composition as a result of CFPE (15, 18). Both studies relied on paired samples (two samples obtained at defined times from the same patient), however while Fodor *et al*, (2012) examined the changes from initiation of treatment for CFPE to completion (15), Carmody *et al* (2013) compared baseline samples, before the onset of respiratory symptoms associated with CFPE to those samples taken when the exacerbation had been confirmed, but prior to the start of treatment (18). A study by Price et al (2013) attempted to address changes in the community over the entirety of a CFPE by collecting a single sample from each of the following periods; baseline, exacerbation, treatment and recovery. However, this study failed to recognise trends in the bacterial community over the course of the disease (19). Although these studies provided insights into changes in community composition, they were limited by the sampling strategy, which did not allow any indication of short term variation within the airway community to be accounted for. These studies highlight the need to examine community changes using multiple samples per clinical period in conjunction with detailed patient metadata.

In order to address the knowledge gap identified by these studies, FLX Titanium 454 pyrosequencing was used to examine changes in the bacterial community dynamics of 12 patients, over the course of one year. During this period ten of the twelve patients were treated for a CFPE, multiple sputum samples were collected throughout the full cycle of CFPE and back to baseline. This full cycle approach has, as yet, not been explored and therefore this study provides novel insight into the changes in the bacterial community

dynamics between disease states. From this, patterns in the bacterial community over time could be identified, while detecting potential biomarkers for clinical exacerbations.

7.2 Materials and methods

7.2.1 Sample collection

This study was undertaken with the local ethical approval from Southampton and South West Hampshire Research Ethics Committee (06/Q1704/24). Sputum samples were collected from 12 adult CF patients attending the Southampton General Hospital, Southampton, UK (Table 7.1). Subjects were selected due to their persistent production of sputum and history of CFPE. All patients were chronically colonised with *Pseudomonas aeruginosa*.

Samples were collected from patients during periods of stability, defined as periods where patients were only receiving maintenance doses of antibiotics. During the study period 10 of the 12 patients were treated for a CFPE. The start and end of CFPE were identified by treating clinicians, and were defined for the purpose of this study as the period of time where patients received clinical intervention in the form of aggressive antibiotic treatment. Decisions to initiate treatment were based on worsening clinical symptoms (20), stabilisation or improvement of these symptoms lead to the termination of this treatment. Samples were retrospectively partitioned into 5 periods; 1a) Stable pre-CFPE, 2) 30 days prior to antibiotic treatment for CFPE, 3) period of time patients were receiving treatment for pulmonary exacerbation, 4) 30 days post treatment for CFPE, 1b) Stable post-CFPE.

Table 7.1 Clinical characteristics for individual patients

Patient	Age (Years)	Gender	CFTR genotype	BMI	Diabetes	CFPE Antibiotics*
1	30	Male	ΔF508/NK	29	No	Ciprofloxacin PO
2	45	Female	ΔF508/NK	18.2	Yes	Colomycin IV + Tobramycin IV
3	47	Male	ΔF508/NK	19.9	Yes	
4	22	Female	ΔF508/ΔF508	18	No	Cirprofloxacin PO, then, Meropenem IV + Amakacin IV
5	55	Male	ΔF508/G58E	23.9	No	Ceftazidime IV + Gentamicin IV
6	21	Female	ΔF508/ΔF508	20.3	No	Ciprofloxacin PO
7	40	Male	ΔF508/ΔF508	19.4	Yes	
8	22	Male	ΔF508/ΔF508	18.4	Yes	Meropenem IV + Colomycin IV
9	17	Female	ΔF508/ΔF508	22.5	No	Ceftazidime IV + Gentamicin IV
10	24	Female	ΔF508/G542X	21	No	Clarithromycin PO
11	20	Male	ΔF508/ΔF508	20.4	No	Ciprofloxacin PO + Metronidazole
12	20	Male	ΔF508/ΔF508	28.5	No	Ceftazidime IV + Gentamicin IV

Abbreviations: BMI, body mass index (kg /m^2); CFTR, cystic fibrosis transmembrane conductance regulator; NK, genotype not known; * Antibiotics administered as intervention for a clinically defined CFPE; PO, Oral; IV, intravenous.

7.2.2 Clinical information

Clinically relevant symptoms were monitored throughout the sampling period. Lung function was assessed using a Koko PeakPro home spirometer (Ferraris Cardiorespiratory, Louisville, CO, USA) at the time of each sample collection. Patients were also required to assess respiratory symptoms; cough, breathlessness, sputum production and general wellbeing, using a visual analogue scale (VAS) scored 0-100, with 0 being the worst and 100 the best.

7.2.3 DNA extraction and Pyrosequencing

Samples were stored at -80°C until processing. A sterile scalpel was used to transfer approximately 250µl of frozen sputum into a 15ml centrifuge tube for a sputum wash to be performed, as described in Chapter 2.3.1. Washed sputum was then treated with propidium monoazide (PMA) using the method described in Chapter 2.3.2 prior to DNA extraction, Chapter 2.3.3

7.2.4 Sample processing and sequence analysis

The primers 338F (5'-ACTCCTACGGGAGGCAGCAG) and 926R (5'-CCGTCAATTCMTTTRAGT) were used to perform bacterial Golay Barcode encoded FLX Titanium amplicon pyrosequencing following the method outlines in Chapter 2.4.1. 16S rRNA gene amplicons were initially generated using a one step PCR of 25 cycles using AccuPrime™ Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA). 454 pyrosequencing using the Lib-L kit was performed at the Wellcome Trust Sanger Institute, Hinxton, UK.

As previously described in Chapter 2.7, resulting data was analysed using the Mothur sequencing analysis platform. The sequence data reported in this paper have been deposited in the European Nucleotide Archive under Study Accession Numbers ERP005251

and ERP007059, and Sample Accession Numbers ERS421603 and ERS551400. The relevant barcode information for each sample is shown in Table A7.1.

7.2.5 Statistical analysis

All statistical analysis was performed in R version 3.1.1(2012-07-10) (21).

Species were partitioned using a distribution abundance distribution, described by Magurran and Henderson (2003) (22). The most persistent and abundant species, those present in more that 75% of the total samples, were described as common, while all others were considered to be rare.

The rate of species turnover between consecutive samples was calculated using the method described by Brown and Kodric-Brown (1977) (23).

Analysis of similarity (ANOSIM) was used to assess the similarity between disease periods, using PAST (version 2.7) (24). This method uses the Bray-Curtis measure of similarity to assess the similarity within periods. Results were compared between periods to allow differences to be assessed. *R* values are on a scale of +1 to -1, values of +1 indicates the most similar samples are within the same period, while -1 indicates the most similar samples are outside the period. Significant *R* values indicate that the community similarities are more similar within the disease period and therefore can be considered significantly different.

Similarity of Percentages (SIMPER) analysis was used to assess the contribution of each species to the observed similarity between disease periods, in order to identify those species that are important in creating the observed pattern of similarity using the Bray-Curtis measure of similarity. SIMPER analyses were performed as previously described by Clarke (1993) (25), using PAST (version2.7) (24).

To analyse species level changes over the five disease periods, outlined above, mixed effect models (GLMMADMB) with negative binomial errors were used as data was found to be over

dispersed (data showed greater variability than would be expected based on a Poisson distribution). For each species the change in abundance, across all patients, was measured using the disease period as the fixed effect and variation between patients was accounted for by including patient as a random effect. The model fits the changes in abundance on the logit scale. The null hypothesis for each species was that there would be no change in species abundance between periods.

7.3 Results

Sputum samples were collected from 12 adult CF patients over the course of 12 months along with complementary clinical information (Table 7.1). The diversity and composition of bacterial communities was assessed using 16S rRNA gene pyrosequencing. A total of 237 samples were sequenced resulting in a total of 386,002 bacterial sequences (mean± standard error/sample 1628±84, *n*=237), comprising 92 genera and 163 distinct Operational taxonomic units (OTUs) classified to species level (Table A7.2).

In order to examine how bacterial community changes were associated with disease state, data was partitioned into distinct clinical periods, in line with those defined and used by Zhao *et al* (2012) (5) and Price *et al* (2013) (19). The data was partitioned into 5 clinical periods; 1a) stable pre-CFPE (*n*=56, 1845±247), 2) 30 days prior to CFPE treatment (*n*=41, 1357±136), 3) treatment for clinical exacerbation (*n*=67, 1643±127), 4) 30 days post CFPE treatment (*n*=32, 1845±151), 1b) stable post-CFPE (*n*=41, 1449±212). In this study the course of a single exacerbation was followed allowing the addition of a fifth category, stable post-CFPE, to investigate the resistance of the bacterial community to perturbations within the system.

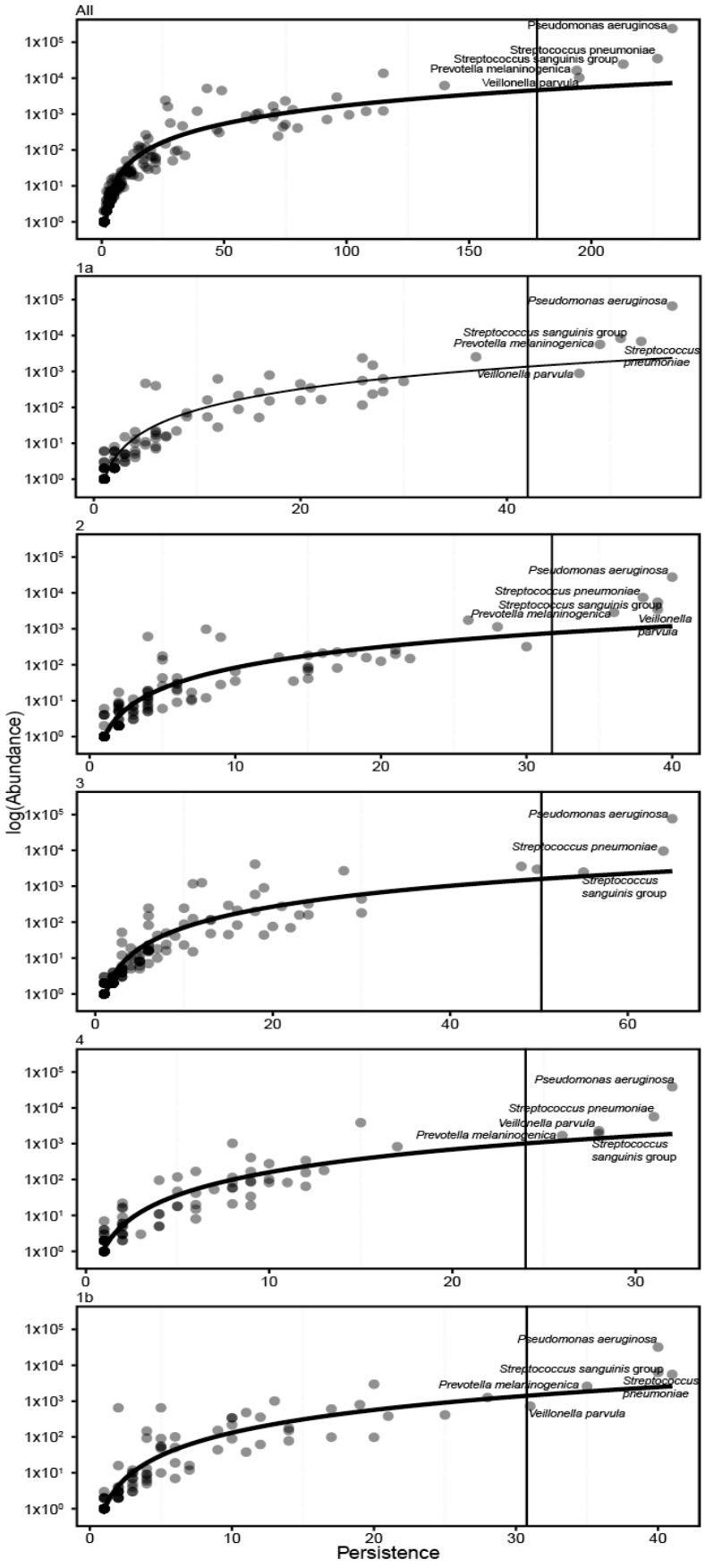
During periods 1a, 2, 4 and 1b, patients received their standard doses of maintenance antibiotics. During period 3, defined as the period of treatment for CFPE, patients were

hospitalised and received increased antibiotic intervention, in the majority of cases through IV antibiotic treatment, see Table 7.1.

Periods 2 and 4 were defined as 30 days pre- and post- antibiotic intervention. These periods were chosen to investigate how the bacterial community changed leading into and out of pulmonary exacerbation. Period 2 was chosen to investigate how the bacterial community changed in the run up to clinical intervention becoming necessary, while period 4 was defined to investigate how the bacterial community recovered after antibiotic intervention. The 30 day period allowed any residual effect of antibiotic treatment to be completely removed from the system prior to the post CFPE baseline period.

7.3.1 Partitioning OTUs

A distribution abundance relationship (DAR) was used to partition the common, most persistent and abundant OTUs, and the rare, transient OTUs found in low abundance. The log OTU abundance was plotted against the sample persistence (the number of samples in which a particular OTU appears) showing a significant positive relationship, see Figure 7.1. This direct relationship between OTU persistence and abundance indicates a coherent metacommunity. OTUs present in more than 75% of the samples were considered common, all remaining OTUs were considered rare, Figure 7.1. Five OTUs were found to be present in more than 75% of the samples and were therefore considered to be common. These common OTUs derived from; *Pseudomonas aeruginosa, Streptococcus pneumoniae, S. sanguinis* group, *Prevotella melaninogenica* and *Veillonella parvula*, were found to add up to over 84% of the total sequences in this study (327,133 sequences). On the other hand, the rare group was comprised of 158 OTUs, and accounted for just over 15% of the total sequences (58,869 sequences).



All samples

1a.
Stable pre-CFPE

2.
30 days prior to antibiotic treatment

3.
Treatment period

4.
30 days post treatment

1b.
Stable post-CFPE

Figure 7.1 The persistence and total abundance of bacteria taxa present in all longitudinal samples and at each of the five partitioned periods.

Common taxa were defined as those that fell within the upper quartile (vertical line), all taxa that fell below the line were considered rare. All) The total from samples collected from the 12 CF patients, irrespective of disease state, within the study (n=237, r^2=0.7, $F_{(1,182)}$=425.7, P<0.001). 1a) Stable pre-CFPE (n=56, r^2=0.8, $F_{(1,106)}$=429.3, P<0.001), 2) 30 days prior to antibiotic treatment (n=41, r^2=0.8, $F_{(1,121)}$=415.1, P<0.001), 3) period of time patients were receiving treatment for pulmonary exacerbation (n=67, r^2=0.7, $F_{(1,140)}$=363.3, P<0.001), 4) 30 days post treatment (n=32, r^2=0.8, $F_{(1,86)}$=316.7, P<0.001), 1b) Stable post-CFPE (n=41, r^2=0.7, $F_{(1,92)}$=221.1, P<0.001)

When samples were divided up by clinical state and partitioned it was found that the *Veillonella parvula* and *Prevotella melaninogenica* OTUs fell below the upper quartile, resulting in those OTUs being considered rare during the treatment period (Figure 7.1, 3). Importantly however, these OTUs were found to return to the common group after the conclusion of the treatment period, Figure 7.1. All other OTUs were consistently partitioned into either the common or rare group.

7.3.2 OTU turnover

The rate of OTU turnover, the number of OTUs eliminated and replaced over time, was assessed based on for each patient over the study period using the Brown and Kordric-Brown (1977) measure of species turnover, Figure 7.2 (23). Changes in the rate of OTU turnover were observed showing more variation in the periods surrounding and during treatment for CFPE.

By partitioning the turnover data into 5 periods it was found that during period 1a and 1b the rate of OTU turnover was relatively consistent, this was particularly evident for patients 3 and 7 who did not experience a CFPE over the study period, Figure 7.2. However, during periods 2, 3 and 4 greater levels of variation in this rate were observed, meaning that no statistically significant trend was observed as changes in turnover rates were not consistent between individual patients.

The turnover rates of the partitioned groups were plotted and displayed on Figure 7.1. This revealed that the rate of turnover of the common OTUs was much lower than the rare category. Most turnover of common OTUs occurred during period 2, 3 and 4 however, this was not the case for all patients. The rate of rare OTU turnover was greater than that observed for the common OTUs and showed high levels of variability. This suggests that the rare group members were driving the observed changes in the rate of community turnover over time.

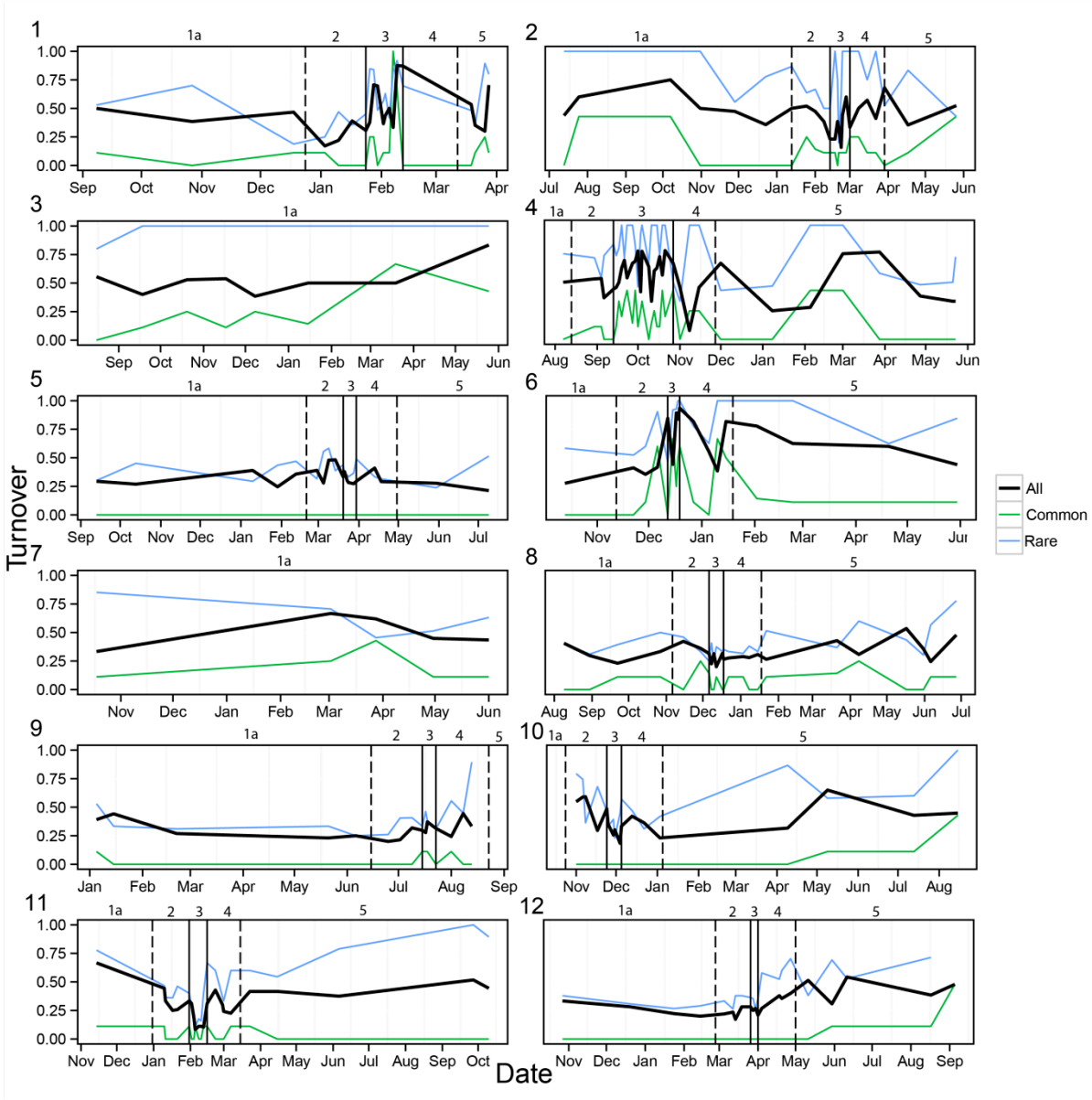


Figure 7.2 Changes in the rate of OTUs turnover for each patient.

Solid vertical lines indicate the start and end of treatment for CFPE, Dashed lines indicate the start and end of the 30 day period either side of the treatment period. 1a; stable pre-CFPE, 2; 30 days prior to CFPE treatment, 3; treatment for clinical exacerbation, 4; 30 days post CFPE treatment, 1b; stable post-CFPE. Black line represents the total community turnover, green represents the common and blue the rare.

7.3.3 Analysis of similarities (ANOSIM)

Analysis of similarity was used to determine how the community composition changed between the 5 disease periods using Bray-Curtis measure of similarity. ANOSIMs were carried out on the whole community and the common and rare groups (Table 7.3). No significant change was observed between disease periods within the whole or common communities. However, significant differences in the community composition of the rare community were seen between; period 1a (stable pre CFPE) and periods 3 (P=0.002) and 4 (P=0.04), period 2 (30 days pre treatment) and period 4 (P=0.009) and 1b (P=0.04), period 4 (30 days post treatment) was also significantly different from 1b (P=0.009). Importantly no significant difference in community composition was observed between the two stable periods, pre and post CFPE (P=0.5). These results support the observations made previously when examining changes in the rate of OTU turnover.

Table 7.2 Comparison of community composition using analysis of similarity (ANOSIM) for each disease period.

ANOSIM was carried out using Bray-Curtis measure of similarity, for the whole, common and rare communities. Matrices below show *R* values in the lower half and *P* values on the upper. Significant *R* values indicate the samples within a group are more similar than would be expected by random chance, therefore the two groups can be considered significantly different.

P values less than 0.05 were considered significant. Significant *P* values are highlight in green.

Whole	1a	2	3	4	1b
1a	*R* *P*	0.21	0.0919	0.4133	0.5622
2	0.01194	*R* *P*	0.2757	0.177	0.342
3	0.01848	0.01222	*R* *P*	0.7783	0.3957
4	0.0003736	0.01658	-0.03389	*R* *P*	0.1993
1b	-0.006949	0.000646	0.002817	0.01431	*R* *P*

Common	1a	2	3	4	1b
1a	*R* *P*	0.1909	0.1428	0.4776	0.5337
2	0.01142	*R* *P*	0.2592	0.2117	0.4807
3	0.01162	0.01288	*R* *P*	0.8338	0.413
4	-0.005413	0.01251	-0.03924	*R* *P*	0.1394
1b	-0.006063	-0.00388	0.001393	0.02168	*R* *P*

Rare	1a	2	3	4	1b
1a	*R* *P*	0.1469	0.0016	0.0361	0.5339
2	0.02272	*R* *P*	0.9613	0.0094	0.0394
3	0.05551	-0.04503	*R* *P*	0.3131	0.2543
4	0.07463	0.07108	0.01644	*R* *P*	0.0086
1b	-0.0041	0.02845	0.01742	0.07378	*R* *P*

7.3.4 Similarity percentage (SIMPER)

Similarity percentage (SIMPER) tables allow the investigation of the contribution of individual OTUs to the community similarity. Using the Bray-Curtis measure of similarity SIMPER tables were made to examine the contributions of OTUs from the whole (Table 7.4.1), common (Table 7.4.2) and rare (Table 7.4.3) communities, between each of the 5 disease periods. Table 7.4.1 indicates that *P. aeruginosa* contributed the most to the whole community similarity between all disease states; this was backed up by the result of the common groups (Table 7.4.2). Of the rare group, *Porphyromonas catoniae* was found to provide the biggest contribution to community similarity in all disease periods, with the exception of the comparison between the CFPE treatment period and 30days post-CFPE, where *Fusobacterium nucleatum* provided the largest contribution.

Table 7.3.1 Similarity of percentages (SIMPER) analysis of the bacterial dissimilarity (Bray-Curtis) of the whole bacterial community between disease states.

Given is the mean % abundance of sequences for each OTU across the periods they were observed to occupy. In addition, the average dissimilarity between periods is given. Percentage contribution is calculated from the mean contribution divided by the mean dissimilarity between periods. A) SIMPER analysis of the whole community between period 1a and 2, B) SIMPER analysis of the whole community between period 2 and 3, C) SIMPER analysis of the whole community between period 3 and 4, D) SIMPER analysis of the whole community between period 4 and 1b, E) SIMPER analysis of the whole community between period 1a and 1b. *Indicates organisms partitioned as common; organisms in bold are considered to be recognised CF pathogens.

A

Taxon	% Mean abundance		Average dissimilarity	Contribution %	Cumulative %
	1a	2			
Pseudomonas aeruginosa*	53.9	49	22.03	37.44	37.44
Streptococcus pneumoniae*	8.22	12.9	7.056	11.99	49.42
Streptococcus sanguinis group*	8.53	9.5	5.603	9.521	58.95
*Prevotella melaninogenica**	9.35	5.12	5.128	8.714	67.66
Porphyromonas catoniae	4.91	2.99	3.142	5.339	73
*Veillonella parvula**	1.21	5.41	2.605	4.426	77.42
Fusobacterium nucleatum	2.03	2.76	2.215	3.763	81.19
Enterobacter cowanii	0.0144	2.51	1.262	2.144	83.33
Stenotrophomonas maltophilia	2.37	0.141	1.225	2.082	85.41
Prevotella enoeca	0.559	0.915	0.7029	1.194	86.61
Prevotella oris	1.22	0.315	0.6859	1.165	87.77
Staphylococcus aureus	0.0172	1.13	0.5699	0.9683	88.74
Bacteroides oleiciplenus	0.784	0.546	0.5613	0.9537	89.7
Megasphaera micronuciformis	0.729	0.352	0.4751	0.8072	90.5

Table 7.3.1. Continued

B

Taxon	% Mean abundance 2	% Mean abundance 3	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa*	49	57.2	21.81	37.33	37.33
Streptococcus pneumoniae*	12.9	12	8.164	13.97	51.3
Streptococcus sanguinis group*	9.5	3.37	4.57	7.823	59.13
*Prevotella melaninogenica**	5.12	4.5	3.452	5.91	65.04
*Veillonella parvula**	5.41	4.07	3.387	5.798	70.84
Enterobacter cowanii	2.51	3.95	3.061	5.239	76.07
Fusobacterium nucleatum	2.76	2.09	2.259	3.867	79.94
Porphyromonas catoniae	2.99	0.89	1.726	2.954	82.9
Staphylococcus aureus	1.13	1.66	1.339	2.293	85.19
Prevotella enoeca	0.915	1.04	0.9239	1.582	86.77
Achromobacter xylosoxidans	0.19	1.03	0.5872	1.005	87.77
Haemophilus parainfluenzae	0.381	0.828	0.546	0.9346	88.71
Barnesiella intestinihominis	0.269	0.844	0.5298	0.9069	89.62
Stenotrophomonas maltophilia	0.141	0.904	0.5051	0.8647	90.48

C

Taxon	% Mean abundance 3	% Mean abundance 4	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa*	57.2	61.9	21.2	38.84	38.84
Streptococcus pneumoniae*	12	9.4	7.706	14.12	52.96
Fusobacterium nucleatum	2.09	6.92	4.206	7.706	60.67
*Prevotella melaninogenica**	4.5	3.53	3.247	5.949	66.62
*Veillonella parvula**	4.07	4.2	2.974	5.45	72.07
Streptococcus sanguinis group*	3.37	3.37	2.319	4.249	76.32
Enterobacter cowanii	3.95	0.0414	1.987	3.64	79.96
Porphyromonas catoniae	0.89	3.15	1.842	3.374	83.33
Prevotella enoeca	1.04	1.84	1.323	2.424	85.76
Staphylococcus aureus	1.66	0.0042	0.8324	1.525	87.28
Achromobacter xylosoxidans	1.03	0.309	0.6343	1.162	88.45
Stenotrophomonas maltophilia	0.904	0.0233	0.4585	0.8402	89.29
Haemophilus parainfluenzae	0.828	0.116	0.449	0.8227	90.11
Barnesiella intestinihominis	0.844	0.0338	0.4308	0.7894	90.9

Table 7.3.1. Continued

D

Taxon	% Mean abundance 4	% Mean abundance 1b	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa*	61.9	53.5	20.4	37.42	37.42
Streptococcus pneumoniae*	9.4	9.45	6.589	12.08	49.5
Fusobacterium nucleatum	6.92	3.94	4.838	8.872	58.38
Streptococcus sanguinis group*	3.37	9.44	4.448	8.157	66.53
*Prevotella melaninogenica**	3.53	4.78	3.238	5.939	72.47
Porphyromonas catoniae	3.15	2	2.154	3.951	76.42
*Veillonella parvula**	4.2	1.3	2.133	3.912	80.34
Prevotella enoeca	1.84	1.69	1.539	2.823	83.16
Stenotrophomonas maltophilia	0.0233	1.79	0.8973	1.646	84.8
Staphylococcus aureus	0.0042	1.76	0.8823	1.618	86.42
Neisseria mucosa	0.155	1.44	0.766	1.405	87.83
Nocardia cyriacigeorgica	0.00775	1.27	0.6387	1.171	89
Prevotella oris	0.598	0.742	0.5766	1.058	90.06
Sneathia sanguinegens	0.706	0.49	0.5459	1.001	91.06

E

Taxon	% Mean abundance 1a	% Mean abundance 1b	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa*	53.9	53.5	21.43	37.7	37.7
Streptococcus pneumoniae*	8.22	9.45	5.889	10.36	48.06
Streptococcus sanguinis group*	8.53	9.44	5.508	9.689	57.75
*Prevotella melaninogenica**	9.35	4.78	5.048	8.879	66.63
Porphyromonas catoniae	4.91	2	2.72	4.785	71.41
Fusobacterium nucleatum	2.03	3.94	2.717	4.779	76.19
Stenotrophomonas maltophilia	2.37	1.79	1.953	3.436	79.63
Prevotella enoeca	0.559	1.69	1.049	1.846	81.47
Neisseria mucosa	0.717	1.44	1.004	1.766	83.24
*Veillonella parvula**	1.21	1.3	0.8938	1.572	84.81
Staphylococcus aureus	0.0172	1.76	0.8881	1.562	86.37
Prevotella oris	1.22	0.742	0.8703	1.531	87.91
Nocardia cyriacigeorgica	0.00321	1.27	0.6369	1.12	89.03
Parvimonas micros	0.333	0.685	0.4852	0.8535	89.88
Bacteroides oleiciplenus	0.784	0.321	0.476	0.8373	90.72

Table 7.3.2 Similarity of percentages (SIMPER) analysis of the bacterial dissimilarity (Bray-Curtis) of the common taxa group between disease states.

Given is the mean % abundance of sequences for each OTU across the periods they were observed to occupy. In addition, the average dissimilarity between periods is given. Percentage contribution is calculated from the mean contribution divided by the mean dissimilarity between periods. A) SIMPER analysis of the common group between period 1a and 2, B) SIMPER analysis of the common group between period 2 and 3, C) SIMPER analysis of the common group between period 3 and 4, D) SIMPER analysis of the common group between period 4 and 1b, E) SIMPER analysis of the common group between period 1a and 1b. Organisms in bold are considered to be recognised CF pathogens.

A

	% Mean abundance		Average	Contribution %	Cumulative %
Taxon	**1a**	**2**	**dissimilarity**		
Pseudomonas aeruginosa	53.9	49	27.55	51.44	51.44
Streptococcus pneumoniae	8.22	12.9	8.933	16.68	68.11
Streptococcus sanguinis group	8.53	9.5	7.075	13.21	81.32
Prevotella melaninogenica	9.35	5.12	6.701	12.51	93.83
Veillonella parvula	1.21	5.41	3.304	6.167	100

B

	% Mean abundance		Average	Contribution %	Cumulative %
Taxon	**2**	**3**	**dissimilarity**		
Pseudomonas aeruginosa	49	57.2	28.05	52.72	52.72
Streptococcus pneumoniae	12.9	12	10.4	19.54	72.26
Streptococcus sanguinis group	9.5	3.37	5.866	11.02	83.28
Veillonella parvula	5.41	4.07	4.461	8.383	91.66
Prevotella melaninogenica	5.12	4.5	4.437	8.338	100

Table 7.3.2. Continued

C

Taxon	% Mean abundance 3	% Mean abundance 4	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa	57.2	61.9	28.3	57.46	57.46
Streptococcus pneumoniae	12	9.4	9.852	20	77.46
Prevotella melaninogenica	4.5	3.53	4.152	8.431	85.89
Veillonella parvula	4.07	4.2	3.919	7.957	93.85
Streptococcus sanguinis group	3.37	3.37	3.03	6.152	100

D

Taxon	% Mean abundance 4	% Mean abundance 1b	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa	61.9	53.5	27.59	56.28	56.28
Streptococcus pneumoniae	9.4	9.45	8.492	17.32	73.61
Streptococcus sanguinis group	3.37	9.44	5.907	12.05	85.66
Prevotella melaninogenica	3.53	4.78	4.211	8.59	94.25
Veillonella parvula	4.2	1.3	2.82	5.753	100

E

Taxon	% Mean abundance 1a	% Mean abundance 1b	Average dissimilarity	Contribution %	Cumulative %
Pseudomonas aeruginosa	53.9	53.5	28.16	54.87	54.87
Streptococcus pneumoniae	8.22	9.45	7.685	14.97	69.84
Streptococcus sanguinis group	8.53	9.44	7.275	14.17	84.02
Prevotella melaninogenica	9.35	4.78	6.892	13.43	97.44
Veillonella parvula	1.21	1.3	1.311	2.555	100

Table 7.3.3 Similarity of percentages (SIMPER) analysis of the bacterial dissimilarity (Bray-Curtis) of the rare taxa group between disease states.

Given is the mean % abundance of sequences for each OTU across the periods they were observed to occupy. In addition, the average dissimilarity between periods is given. Percentage contribution is calculated from the mean contribution divided by the mean dissimilarity between periods. A) SIMPER analysis of the rare group between period 1a and 2, B) SIMPER analysis of the rare group between period 2 and 3, C) SIMPER analysis of the rare group between period 3 and 4, D) SIMPER analysis of the rare group between period 4 and 1b, E) SIMPER analysis of the rare group between period 1a and 1b. Organisms in bold are considered to be recognised CF pathogens

A

Taxon	% Mean abundance		Average dissimilarity	Contribution %	Cumulative %
	1a	2			
Porphyromonas catoniae	4.91	2.99	17.67	19.48	19.48
Fusobacterium nucleatum	2.03	2.76	6.669	7.352	26.84
Stenotrophomonas maltophilia	2.37	0.141	5.237	5.773	32.61
Enterobacter cowanii	0.0144	2.51	4.923	5.428	38.04
Prevotella oris	1.22	0.315	3.649	4.023	42.06
Bacteroides oleiciplenus	0.784	0.546	3.486	3.843	45.9
Megasphaera micronuciformis	0.729	0.352	3.429	3.78	49.68
Haemophilus parainfluenzae	0.0138	0.381	2.926	3.226	52.91
Gemella sanguinis	0.183	0.489	2.905	3.203	56.11
Capnocytophaga sputigena	0.123	0.506	2.786	3.071	59.18
Prevotella maculosa	0.74	0.255	2.584	2.849	62.03
Prevotella pallens	0.471	0.401	2.427	2.676	64.71
Staphylococcus aureus	0.0172	1.13	2.413	2.66	67.37
Prevotella nanceiensis	0.305	0.349	2.09	2.304	69.67
Neisseria mucosa	0.717	0.177	2.064	2.275	71.95
Granulicatella adiacens	0.308	0.199	2.049	2.259	74.21
Capnocytophaga granulosa	0.17	0.409	2.043	2.253	76.46
Prevotella enoeca	0.559	0.915	1.722	1.898	78.36
Oribacterium sinus	0.257	0.139	1.593	1.757	80.12
Achromobacter xylosoxidans	0.028	0.19	1.323	1.459	81.57
Prevotella loescheii	0.283	0.123	1.251	1.379	82.95
Nocardia cyriacigeorgica	0.00321	0.181	1.243	1.37	84.32
Parvimonas micros	0.333	0.521	1.117	1.232	85.56
Porphyromonas endodontalis	0.416	0.0559	1.026	1.132	86.69
Lachnoanaerobaculum orale	0.218	0.149	0.9714	1.071	87.76
Actinomyces odontolyticus	0.282	0.0664	0.9427	1.039	88.8
Barnesiella intestinihominis	0.00679	0.269	0.7411	0.8171	89.62
Prevotella nigrescens	0.242	0.112	0.7262	0.8007	90.42

Table 7.3.3 Continued

B) SIMPER analysis of the rare group between period 2 (30 days prior to antibiotic treatment for CFPE) and 3 (Treatment)

B

Taxon	% Mean abundance 2	% Mean abundance 3	Average dissimilarity	Contribution %	Cumulative %
Porphyromonas catoniae	2.99	0.89	11.28	12.19	12.19
Enterobacter cowanii	2.51	3.95	9.412	10.17	22.36
Fusobacterium nucleatum	2.76	2.09	6.643	7.177	29.53
Haemophilus parainfluenzae	0.381	0.828	4.675	5.051	34.58
Staphylococcus aureus	1.13	1.66	4.652	5.026	39.61
Achromobacter xylosoxidans	0.19	1.03	4.033	4.357	43.97
Prevotella pallens	0.401	0.559	3.035	3.279	47.25
Prevotella enoeca	0.915	1.04	2.904	3.137	50.38
Capnocytophaga sputigena	0.506	0.0623	2.851	3.08	53.46
Gemella sanguinis	0.489	0.147	2.801	3.026	56.49
Barnesiella intestinihominis	0.269	0.844	2.748	2.968	59.46
Capnocytophaga granulosa	0.409	0.263	2.61	2.819	62.28
Prevotella nanceiensis	0.349	0.404	2.494	2.694	64.97
Megasphaera micronuciformis	0.352	0.196	2.404	2.598	67.57
Bacteroides oleiciplenus	0.546	0.183	2.34	2.528	70.1
Prevotella oris	0.315	0.347	2.176	2.351	72.45
Stenotrophomonas maltophilia	0.141	0.904	2.135	2.306	74.75
Neisseria mucosa	0.177	0.228	1.901	2.054	76.81
Prevotella maculosa	0.255	0.294	1.856	2.005	78.81
Granulicatella adiacens	0.199	0.0668	1.5	1.621	80.43
Nocardia cyriacigeorgica	0.181	0.00366	1.283	1.386	81.82
Oribacterium sinus	0.139	0.0538	1.263	1.365	83.18
Lachnoanaerobaculum orale	0.149	0.252	1.144	1.236	84.42
Prevotella loescheii	0.123	0.2	0.9566	1.033	85.45
Parvimonas micros	0.521	0.107	0.7968	0.8607	86.31
Prevotella nigrescens	0.112	0.114	0.6742	0.7283	87.04
Clostridium cellobioparum	0.017	0.192	0.6735	0.7276	87.77
Atopobium parvulum	0.0605	0.0537	0.5502	0.5944	88.36
Actinomyces odontolyticus	0.0664	0.0785	0.5133	0.5545	88.92
Haemophilus influenzae	0.0569	0.00545	0.5119	0.553	89.47
Clostridium lavalense	0.066	0.125	0.4467	0.4825	89.95
Sneathia sanguinegens	0.177	0.163	0.4276	0.4619	90.41

Table 7.3.3 Continued

C) SIMPER analysis of the rare group between period 3 (treatment) and 4 (30 days post treatment for CFPE)

C

Taxon	% Mean abundance 3	% Mean abundance 4	Average dissimilarity	Contribution %	Cumulative %
Fusobacterium nucleatum	2.09	6.92	11.65	12.39	12.39
Porphyromonas catoniae	0.89	3.15	11.09	11.79	24.19
Achromobacter xylosoxidans	1.03	0.309	5.699	6.062	30.25
Enterobacter cowanii	3.95	0.0414	5.65	6.01	36.26
Haemophilus parainfluenzae	0.828	0.116	4.529	4.818	41.08
Prevotella enoeca	1.04	1.84	4.412	4.693	45.77
Barnesiella intestinihominis	0.844	0.0338	3.267	3.475	49.25
Capnocytophaga granulosa	0.263	0.336	3.238	3.444	52.69
Prevotella oris	0.347	0.598	2.748	2.923	55.61
Prevotella pallens	0.559	0.122	2.705	2.877	58.49
Staphylococcus aureus	1.66	0.0042	2.653	2.822	61.31
Neisseria mucosa	0.228	0.155	2.499	2.658	63.97
Prevotella maculosa	0.294	0.242	2.438	2.594	66.56
Capnocytophaga sputigena	0.0623	0.518	2.262	2.406	68.97
Prevotella nanceiensis	0.404	0.167	2.183	2.323	71.29
Gemella sanguinis	0.147	0.248	2.093	2.226	73.52
Stenotrophomonas maltophilia	0.904	0.0233	1.942	2.066	75.58
Megasphaera micronuciformis	0.196	0.156	1.787	1.901	77.49
Lachnoanaerobaculum orale	0.252	0.127	1.526	1.623	79.11
Granulicatella adiacens	0.0668	0.131	1.485	1.58	80.69
Bacteroides oleiciplenus	0.183	0.241	1.392	1.48	82.17
Sneathia sanguinegens	0.163	0.706	1.102	1.172	83.34
Prevotella loescheii	0.2	0.0527	1.034	1.1	84.44
Prevotella nigrescens	0.114	0.154	1.014	1.079	85.52
Oribacterium sinus	0.0538	0.0814	0.6996	0.7442	86.26
Atopobium parvulum	0.0537	0.0531	0.6717	0.7145	86.98
Actinomyces odontolyticus	0.0785	0.079	0.6314	0.6716	87.65
Clostridium cellobioparum	0.192	0	0.6281	0.6682	88.32
Haemophilus influenzae	0.00545	0.0212	0.5675	0.6036	88.92
Parvimonas micros	0.107	0.295	0.4805	0.5111	89.43
Pseudomonas fragi	0.00966	0	0.38	0.4042	89.84
Prevotella oralis	0.137	0.0829	0.377	0.401	90.24
Clostridium lavalense	0.125	0	0.3739	0.3977	90.64

Table 7.3.3 Continued

D) SIMPER analysis of the rare group between period 4 (30 days post treatment for CFPE) and 1b (Stable post-CFPE)

D

Taxon	% Mean abundance		Average dissimilarity	Contribution %	Cumulative %
	4	1b			
Porphyromonas catoniae	3.15	2	14.6	15.95	15.95
Fusobacterium nucleatum	6.92	3.94	13.11	14.32	30.27
Prevotella enoeca	1.84	1.69	4.462	4.873	35.14
Achromobacter xylosoxidans	0.309	0.338	4.239	4.63	39.77
Nocardia cyriacigeorgica	0.00775	1.27	4.105	4.483	44.25
Granulicatella adiacens	0.131	0.446	3.765	4.112	48.37
Stenotrophomonas maltophilia	0.0233	1.79	3.392	3.704	52.07
Gemella sanguinis	0.248	0.54	3.378	3.689	55.76
Prevotella oris	0.598	0.742	3.094	3.379	59.14
Neisseria mucosa	0.155	1.44	3.035	3.315	62.45
Oribacterium sinus	0.0814	0.358	2.458	2.685	65.14
Staphylococcus aureus	0.0042	1.76	2.185	2.387	67.53
Bacteroides oleiciplenus	0.241	0.321	2.16	2.359	69.88
Capnocytophaga sputigena	0.518	0.0937	2.059	2.249	72.13
Prevotella maculosa	0.242	0.225	1.97	2.152	74.29
Capnocytophaga granulosa	0.336	0.0911	1.794	1.959	76.24
Haemophilus parainfluenzae	0.116	0.295	1.669	1.822	78.07
Actinomyces odontolyticus	0.079	0.247	1.662	1.815	79.88
Sneathia sanguinegens	0.706	0.49	1.506	1.644	81.53
Prevotella nanceiensis	0.167	0.221	1.381	1.508	83.04
Megasphaera micronuciformis	0.156	0.238	1.378	1.505	84.54
Parvimonas micros	0.295	0.685	1.197	1.307	85.85
Atopobium parvulum	0.0531	0.102	1.119	1.222	87.07
Prevotella nigrescens	0.154	0.493	1.086	1.186	88.26
Prevotella denticola	0.0059	0.135	0.9884	1.08	89.33
Lachnoanaerobaculum orale	0.127	0.122	0.9742	1.064	90.4

Table 7.3.3 Continued

E) SIMPER analysis of the rare group between period 1a (Stable pre-CFPE) and 1b (Stable post-CFPE)

E

Taxon	% Mean abundance 1a	% Mean abundance 1b	Average dissimilarity	Contribution %	Cumulative %
Porphyromonas catoniae	4.91	2	16.61	18.35	18.35
Fusobacterium nucleatum	2.03	3.94	8.157	9.008	27.36
Stenotrophomonas maltophilia	2.37	1.79	7.242	7.998	35.35
Prevotella oris	1.22	0.742	4.451	4.915	40.27
Granulicatella adiacens	0.308	0.446	3.744	4.135	44.4
Nocardia cyriacigeorgica	0.00321	1.27	3.696	4.082	48.49
Neisseria mucosa	0.717	1.44	3.25	3.589	52.07
Bacteroides oleiciplenus	0.784	0.321	3.107	3.431	55.51
Gemella sanguinis	0.183	0.54	3.045	3.362	58.87
Prevotella enoeca	0.559	1.69	3.042	3.36	62.23
Achromobacter xylosoxidans	0.028	0.338	2.668	2.946	65.17
Prevotella maculosa	0.74	0.225	2.551	2.818	67.99
Oribacterium sinus	0.257	0.358	2.506	2.768	70.76
Megasphaera micronuciformis	0.729	0.238	2.393	2.643	73.4
Staphylococcus aureus	0.0172	1.76	2.237	2.47	75.87
Actinomyces odontolyticus	0.282	0.247	1.832	2.023	77.89
Prevotella pallens	0.471	0.156	1.654	1.827	79.72
Prevotella nanceiensis	0.305	0.221	1.431	1.58	81.3
Parvimonas micros	0.333	0.685	1.307	1.443	82.74
Atopobium parvulum	0.0627	0.102	1.167	1.288	84.03
Porphyromonas endodontalis	0.416	0.072	1.108	1.223	85.26
Prevotella nigrescens	0.242	0.493	1.081	1.193	86.45
Prevotella denticola	0.0533	0.135	1.036	1.145	87.59
Capnocytophaga granulosa	0.17	0.0911	0.977	1.079	88.67
Lachnoanaerobaculum orale	0.218	0.122	0.9638	1.064	89.74
Prevotella loescheii	0.283	0.00487	0.8468	0.9352	90.67

7.3.5 Mixed effect models

Despite observing no significant changes in the core group by disease period (Table 7.3), SIMPER analysis revealed possible changes in the percentage contribution of individual OTUs within the core group. To investigate how the abundance of individual OTUs changed between disease periods mixed effect models were used. These models allow the inclusion of both fixed (disease period) and random effects (Patient) allowing the variation between patients to be accounted for within the model. The five taxa partitioned into the common group were modelled to examine how the relative abundance of these changed over the 5 disease periods, Figure 7.3.

As shown in Figure 7.3, there was no significant difference in the proportional abundance of *P. aeruginosa* prior, or during intervention for CFPE. A significant increase was observed in the proportional abundance of *P. aeruginosa* (100%) after the treatment period (period 4) however, this returned to baseline after the 30 day period. No significant change in the proportional abundance of *S. pneumonia* was observed over the five disease periods. A significant decrease in the proportional abundance of *S. sanguinis* group (82%) and *P. melaninogenica* (59%) was observed during the treatment period (3). However, *V. parvula* was found to show a significant increase in proportional abundance prior to the start of treatment (264%). The proportional abundance remained significantly higher for period 3 (129%) and 4 (400%), with a slight decrease being observed during treatment.

SIMPER analysis indicated the percentage contribution of *P. catoniae*, the most proportionally abundant of the rare group, decreased during the treatment period. This was also supported by the results of the mixed effect modelling, which revealed the proportional abundance of *P. catoniae* significantly decreased (60%) during the treatment period, before recovery leading into the post CFPE stable period.

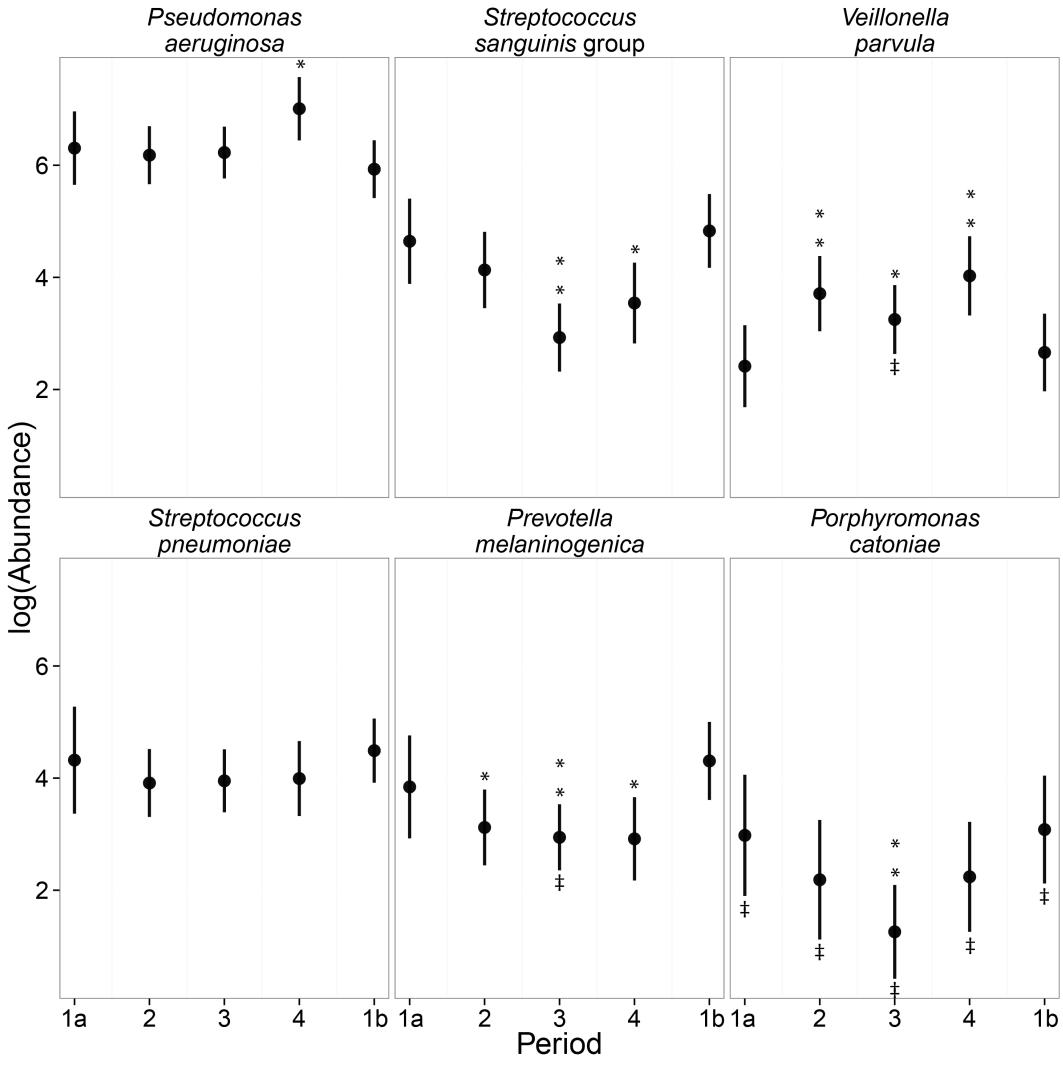


Figure 7.3 Trends in the relative abundance of bacterial taxa with change in disease period.

Parameters are extracted from mixed-effects models based on 237 samples from 12 patients. 1a) Stable pre-CFPE, 2) 30 days prior to antibiotic treatment, 3) period of time patients were receiving treatment for pulmonary exacerbation, 4) 30 days post treatment, 1b) Stable post-CFPE. ** $P<0.001$, *=$P<0.05$. ‡ indicates OTU was partitioned as rare.

7.4 Discussion

Spatial studies have informed our understanding of the bacterial community within the CF lung, revealing a complex and highly variable system. However, in order to understand the progressive nature of CF lung disease and how this relates to changes in bacterial community dynamics, it is important to investigate how the community changes over time. Several recent studies have been published that attempt to understand the bacterial community dynamics within the CF lung (5, 15-19). While these investigations have provided valuable information, they have several limitations, typically, they have relied on point samples to be representative of a given disease state (e.g. baseline or exacerbation), which fails to account for variation in microbiota structure and composition within a disease period. By collecting samples and associated metadata at more frequent intervals this study was able to not only account for variation within the bacterial community over time but also investigate how this relates to changes in disease state.

Despite efforts to investigate change in the bacterial community due to CFPE, the findings of studies by Fodor *et al* (2012) Carmody *et al* (2013) Price *et al* (2013) and Zhao *et al* (2012), were limited (5, 15, 18, 19). Fodor *et al* (2012) and Carmody *et al* (2013) compared paired samples from patients, leading into or during treatment for CFPE (15, 18). While Fodor *et al* (2012) showed no significant change in the bacterial density or composition from the start to the end of treatment for CFPE (15), Carmody *et al* (2013) revealed that changes in the bacterial community, from baseline leading into CFPE was highly variable depending on bacterial diversity and composition (18). While these studies are informative they have been shown to have inherent limitations. Carmody *et al* (2013) state the limitations of their study, including the frequency of the sample collection, which failed to allow fluctuations in the airway microbiota to be accounted for (18). On top of this, both these studies fail to investigate bacterial changes across an entire CFPE, concentrating instead on the treatment period and lead in respectively, they fail to address the issues of how treatment effects the bacterial community and if the community returns to its baseline state post treatment.

Studies by Zhao *et al* (2012) and Price *et al* (2013), aimed to examine changes in the bacterial community over the course of an exacerbation using the "BETR" classification; baseline (B), exacerbation prior to treatment (E), treatment (T) and into recovery (R) (5, 19). While Zhao et al, 2012, used these clearly defined parameters of each disease period (5), Price *et al* (2013) failed to clearly define the recovery period, simply collecting samples at the next routine quarterly visit (19). These studies provided a more complete picture of changes in the bacterial community over the course of CFPE. However, Zhao *et al* (2012) failed to follow a single exacerbation from start to finish for any given patient and therefore revealed little about specific changes in the bacterial community (5). Whereas Price *et al* (2013) examined changes in community composition over a single exacerbation, revealing a stable bacterial community this study. Even so, this work neglected to account for the fluctuations in the community identified previously (19).

In order to address how the bacterial community changes from baseline and across the full cycle of CFPE back to baseline, which has not been previously studied, samples were collected from CF patients over the course of a year. This collection regimen included several samples collected when patients were considered to be stable as well as multiple samples that were collected leading into, during and out of a CFPE, giving a more complete picture of the exacerbation event. The collection of multiple samples allows a more complete assessment of short term variation within the community, therefore allowing true shifts in the bacterial community composition to be observed.

It has been well established in previous studies that the metacommunity within the CF lung shows a distinct divide between common and locally abundant species and those that are rare or transient (3, 26). As explained in Chapter 3.4, such studies used a Poisson distribution to establish the core-satellite groups within the community. However, due to the tendency for this analysis to assign some species with low persistence but uniform abundance as core, a more intuitive approach based on the positive persistence abundance relationship was employed. This approach partitioned only the species appearing in more than 75% of the samples into the common group. Partitioning all samples revealed five

species common OTUs; *Pseudomonas aeruginosa, Streptococcus pneumonia, S. sanguinis* group, *Prevotella melaninogenica* and *Veillonella parvula*. This common group was found to account for the majority of the sequence abundance. However, the majority of diversity was accounted for by the rare group. Interestingly, over the five clinical periods the common group was found to be relatively stable, with only *V. parvula* and *P. melaninogenica* falling into the rare group during the treatment period. However, none of the members of the rare group were found to move into the core.

As expected, bacterial community turnover within the lung was never completely stable however, rates of OTU turnover were found to be more consistent over the stable periods, and show more variation surrounding, and throughout the treatment for CFPE. Changes in species richness were hypothesised to drive the changes in the rate of turnover. To explore this further the turnover rates for the common and rare taxa were calculated. While the common OTUs were found to be conserved over the study period, the much more diverse, rare group was found to be highly variable, driving the turnover rates observed in the whole community.

The Bray-Curtis measure of similarity was used to compare changes in community composition over the study period, allowing a quantitative measure of similarity to be calculated. Over the five clinical periods, no change in similarity was observed when either the whole community or common group were examined. However, significant differences were seen within the rare group. Importantly these changes occurred leading into, during and out of CFPE. However, no significant differences in similarity between the stable periods, pre and post CFPE were observed. These findings underline the importance of partitioning the data, as the contribution of the rare taxa to community composition would be neglected without this distinction. If partitioning had been carried out by Stressmann *et al* (2012), more variation in the community composition may have been observed (16).

Failure of the common group to show any significant change over the study period indicates resistance of that group to perturbations within the lung. Resistance of ecological

communities to change is defined as the degree to which a microbial community can resist change as a result of community perturbations (27). The abundance of these common taxa results in this pattern of resistance being reflected in the whole community, and as a result masks changes in the rare group. While the common group shows resistance to perturbations the data suggests that the rare group shows resilience. Resilience of ecological communities to change is defined as the rate at which the microbial community composition returns to the baseline after perturbation (27). This resilience indicates that over the short term the bacterial community within the lung can, despite the influence of CFPE and interventions, recover to the previous baseline community composition. These findings are reflected in the results observed when examining rates of taxa turnover.

There are two plausible explanations of how the rare taxa can return to baseline after the initiation of CFPE and treatment. The first is that species recolonise the lung through immigration from the upper airways and/or oral cavity, the second is that populations of rare taxa are not totally eradicated by antibiotic intervention and have fallen below the detection threshold, thereby allowing the species to re-establish after the termination of antibiotic intervention. It is probable that both of these explanations will be the case for different rare taxa.

From the community level analysis it is clear that the common group shows high resistance to community perturbations. However, the changes in individual species influence on percentage dissimilarity indicated that subtle changes in the abundance of individual species could be occurring. In order to investigate this mixed effect models, with a negative binomial error structure, were used to examine how individual species from the common group were affected by perturbation at the population level. The most persistent of the rare species group was included also in this analysis due to their high abundance, observed in the persistence abundance plots, along with the high percentage contribution to community similarity, shown in the SIMPER analysis.

Infection with *P. aeruginosa* has been highlighted by numerous studies to be a key factor in poor clinical outcomes and is one of the main targets for treatment. It is therefore surprising that its abundance remained high across all clinical periods, even showing a significant increase in abundance after the treatment period. While the relative abundance from high throughput sequencing data of a given taxa may not be considered strictly quantitative, Price *et al* (2013) clearly demonstrated a highly significant correlation for *P. aeruginosa* between the results of qPCR analysis and sequencing data (19). Therefore, coupled with the PMA treatment, ruling out artefacts of cell death as a result of antibiotic intervention, confidence can be taken that this represents a true reflection of the *P. aeruginosa* population over all disease periods. One possible explanation for the increase in abundance of *P. aeruginosa* observed after the conclusion of treatment is that antibiotic treatment has lead to a reduction in the abundance of other taxa, thereby allowing *P. aeruginosa* to take advantage of newly available niche space. Even so, the abundance in *P. aeruginosa* returns to baseline after the 30 day post treatment period, due to the resilience of members of the rare group.

Similarly, no change in the abundance of *Streptococcus pneumoniae* was observed over the disease period. However, members of the *Streptococcus sanguinis* group were found to significantly decrease during the treatment period. This abundance only returned to baseline after the 30 day post treatment period, suggesting that the initiation of treatment is having an effect on the abundance of the members of the *Streptococcus sanguinis* group but that the group is resilient and is able to re-establish and return to previous baseline abundance.

This resilience of individual taxa to CFPE is also observed when examining the changes in *Porphyromonas catoniae, Prevotella melaninogenica* and *Veillonella parvula*. However, while the abundance of *Porphyromonas catoniae* simply decreases significantly during treatment and then returns to baseline levels, the abundance of *Prevotella melaninogenica* and *V. parvula* significantly changed prior to the initiation of intervention. During the lead in to exacerbation, a significant decrease in the abundance of *Prevotella melaninogenica* was observed, while a significant increase was seen in *V. parvula*. While both these taxa are resilient and return to baseline during the 30 day period after the conclusion of treatment

period, the significant changes observed prior to the start of exacerbation makes them potential biomarkers of disease state.

The lack of a generally applicable definition of CFPE is a major problem for the timely treatment of exacerbations. Work by Sanders *et al* (2010) has already established that in many cases patients suffering from CFPE will fail to return to baseline lung function following treatment however, earlier identification, and therefore treatment, can improve patient outcomes and increase the likelihood of lung function recovery (28). This failure to recover lung function after CFPE, despite the return of the bacterial community to baseline, may be a result to incremental lung damage due to inflammatory responses and not be associated directly with the bacterial community. The advances in DNA based technologies has promoted their use in molecular diagnostics allowing disease associated biomarkers to be used to identify, predict and monitor changes in infections (29). The identification of biomarkers for CFPE could be a potentially important route to improve the identification and therefore outcomes for CF patients.

Two potential biomarkers for disease state, *Prevotella melaninogenica* and *Veillonella parvula,* were identified due to their significant change from baseline prior to the start of antibiotic intervention. While both these species may be useful, the large (264%) increase in the abundance of *V. parvula* observed in this study make it the more promising of the two species. More detailed work into this area is required to examine the potential of these species for use as a diagnostic tool. Development of targeted qPCR based analysis is required to check the efficacy of these species as a molecular diagnostic tool for detecting the onset of CFPE.

Here, examination of changes in the bacterial community across the entire span of CFPE has revealed valuable insight into how the bacterial community reacts to perturbations within the lung. While the common species are found to be resistant, the rare are resilient to CFPE perturbations. Examination at the common species group level has not only revealed potentially two novel biomarkers for changes in disease state but has also revealed that key

recognised CF pathogen *P. aeruginosa* is, in essence, unaffected by a change in disease state.

While the resistance and resilience of the common and rare group is of concern with regards to treatment, of particular concern to clinicians will be the lack of substantial change in *P. aeruginosa*. As antibiotics administered in response to the worsening symptoms associated with CFPE are predominantly targeted at *P. aeruginosa*. These findings suggest that clinicians may need to re-examine intervention regimens. From a fundamental perspective the bacterial community may not be the whole story. The wider microbial community needs to be considered through investigations into the roles of bacteria, fungi and respiratory viruses, alongside markers of host immune response. Studies of this nature could reveal the underlying cause or causes of CFPE. It may well be the case that there is no consistent factor (infection or host) across CFPE, within or between patients.

7.5 References

1. **Dodge JA, Lewis PA, Stanton M, Wilsher J.** 2007. Cystic fibrosis mortality and survival in the UK: 1947-2003. Eur Respir J **29:**522-526.
2. **Rogers GB, Hart CA, Mason JR, Hughes M, Walshaw MJ, Bruce KD.** 2003. Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling. Journal of Clinical Microbiology **41:**3548-3558.
3. **van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD.** 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. Isme J **5:**780-791.
4. **Cox MJ, Allgaier M, Taylor B, Baek MS, Huang YJ, Daly RA, Karaoz U, Andersen GL, Brown R, Fujimura KE, Wu B, Tran D, Koff J, Kleinhenz ME, Nielson D, Brodie EL, Lynch SV.** 2010. Airway Microbiota and Pathogen Abundance in Age-Stratified Cystic Fibrosis Patients. Plos One **5**.
5. **Zhao JC, Schloss PD, Kalikin LM, Carmody LA, Foster BK, Petrosino JF, Cavalcoli JD, VanDevanter DR, Murray S, Li JZ, Young VB, LiPuma JJ.** 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. P Natl Acad Sci USA **109:**5809-5814.
6. **Tunney MM, Klem ER, Fodor AA, Gilpin DF, Moriarty TF, McGrath SJ, Muhlebach MS, Boucher RC, Cardwell C, Doering G, Elborn JS, Wolfgang MC.** 2011. Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. Thorax **66:**579-584.
7. **Flume PA, Mogayzel PJ, Jr., Robinson KA, Goss CH, Rosenblatt RL, Kuhn RJ, Marshall BC, Clinical Practice Guidelines for Pulmonary Therapies C.** 2009. Cystic fibrosis pulmonary guidelines: treatment of pulmonary exacerbations. Am J Respir Crit Care Med **180:**802-808.
8. **Sibley CD, Parkins MD, Rabin HR, Duan K, Norgaard JC, Surette MG.** 2008. A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. P Natl Acad Sci USA **105:**15070-15075.
9. **VanDevanter DR, Elkin EP, Pasta DJ, Morgan WJ, Konstan MW.** Changing thresholds and incidence of antibiotic treatment of cystic fibrosis pulmonary exacerbations, 1995–2005. Journal of Cystic Fibrosis.
10. **Smyth A, Elborn JS.** 2008. Exacerbations in cystic fibrosis: 3 · Management. Thorax **63:**180-184.
11. **Rosenfeld M, Emerson J, Williams-Warren J, Pepe M, Smith A, Montgomery AB, Ramsey B.** 2001. Defining a pulmonary exacerbation in cystic fibrosis. The Journal of Pediatrics **139:**359-365.
12. **Sanders DB, Bittner RCL, Rosenfeld M, Hoffman LR, Redding GJ, Goss CH.** 2010. Failure to Recover to Baseline Pulmonary Function after Cystic Fibrosis Pulmonary Exacerbation. Am J Resp Crit Care **182:**627-632.
13. **Guss AM, Roeselers G, Newton ILG, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM.** 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. Isme J **5:**20-29.
14. **Klepac-Ceraj V, Lemon KP, Martin TR, Allgaier M, Kembel SW, Knapp AA, Lory S, Brodie EL, Lynch SV, Bohannan BJM, Green JL, Maurer BA, Kolter R.** 2010. Relationship between cystic fibrosis respiratory tract bacterial communities and

age, genotype, antibiotics and Pseudomonas aeruginosa. Environ Microbiol **12:**1293-1303.

15. **Fodor AA, Klem ER, Gilpin DF, Elborn JS, Boucher RC, Tunney MM, Wolfgang MC.** 2012. The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. Plos One **7**.
16. **Stressmann FA, Rogers GB, van der Gast CJ, Marsh P, Vermeer LS, Carroll MP, Hoffman L, Daniels TWV, Patel N, Forbes B, Bruce KD.** 2012. Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. Thorax **67:**867-873.
17. **Daniels TWV, Rogers GB, Stressmann FA, van der Gast CJ, Bruce KD, Jones GR, Connett GJ, Legg JP, Carroll MP.** 2013. Impact of antibiotic treatment for pulmonary exacerbations on bacterial diversity in cystic fibrosis. Journal of Cystic Fibrosis **12:**22-28.
18. **Carmody LA, Zhao J, Schloss PD, Petrosino JF, Murray S, Young VB, Li JZ, LiPuma JJ.** 2013. Changes in cystic fibrosis airway microbiota at pulmonary exacerbation. Ann Am Thorac Soc **10:**179-187.
19. **Price KE, Hampton TH, Gifford AH, Dolben EL, Hogan DA, Morrison HG, Sogin ML, O'Toole GA.** 2013. Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. Microbiome **1:**27.
20. **Fuchs HJ, Borowitz DS, Christiansen DH, Morris EM, Nash ML, Ramsey BW, Rosenstein BJ, Smith AL, Wohl ME.** 1994. Effect of Aerosolized Recombinant Human Dnase on Exacerbations of Respiratory Symptoms and on Pulmonary-Function in Patients with Cystic-Fibrosis. New Engl J Med **331:**637-642.
21. **R, Core, Team.** 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
22. **Magurran AE, Henderson PA.** 2003. Explaining the excess of rare species in natural species abundance distributions. Nature **422:**714-716.
23. **Brown JH, Kodricbrown A.** 1977. Turnover Rates in Insular Biogeography - Effect of Immigration on Extinction. Ecology **58:**445-449.
24. **Hammer Ø, Harper, D.A.T., Ryan, P.D.** 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica. **4:**9pp.
25. **Clarke KR.** 1993. Nonparametric Multivariate Analyses of Changes in Community Structure. Aust J Ecol **18:**117-143.
26. **van der Gast CJ, Cuthbertson L, Rogers GB, Pope C, Marsh RL, Redding GJ, Bruce KD, Chang AB, Hoffman LR.** 2014. Three Clinically Distinct Chronic Pediatric Airway Infections Share a Common Core Microbiota. Annals of the American Thoracic Society.
27. **Allison SD, Martiny JBH.** 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences **105:**11512-11519.
28. **Sanders DB, Hoffman LR, Emerson J, Gibson RL, Rosenfeld M, Redding GJ, Goss CH.** 2010. Return of FEV(1) After Pulmonary Exacerbation in Children With Cystic Fibrosis. Pediatr Pulm **45:**127-134.
29. **Pattison SH, Rogers GB, Crockard M, Elborn JS, Tunney MM.** 2013. Molecular detection of CF lung pathogens: Current status and future potential. Journal of Cystic Fibrosis **12:**194-205.

7.6 Appendix

Table A7.1 The raw sequence data reported in this Chapter have been deposited in the European Nucleotide Archive short read database

Study Accession Numbers ERP005251 and ERP007059, and Sample Accession Numbers ERS421603 and ERS551400. A list of barcodes used and their associated samples are shown below.

Barcode	Sputum ID	Sample No	Patient	Date	Exacerbation
CAACTCATCGTA	01_006	6	1	14-Jun-06	1
TACTACATGGTC	01_031	31	1	08-Sep-06	1
ATGACCATCGTG	01_055	55	1	27-Oct-06	1
AGAGTCCTGAGC	01_075	75	1	18-Dec-06	1
ACTATTGTCACG	01_078	78	1	03-Jan-07	2
AGAACACGTCTC	01_081	81	1	10-Jan-07	2
AGCACACCTACA	01_084	84	1	17-Jan-07	2
AGCTCTCAGAGG	01_087	87	1	24-Jan-07	2
AGTCTCGCATAT	01_089	89	1	26-Jan-07	3
ATCGATCTGTGG	01_091	91	1	28-Jan-07	3
ATAGGCGATCTC	01_093	93	1	30-Jan-07	3
AGTTCTACGTCA	01_095	95	1	02-Feb-07	3
TAGTGCTGCGTA	01_096	96	1	03-Feb-07	3
CAAGATCGACTC	01_097	97	1	05-Feb-07	3
ATGTGCACGACT	01_099	99	1	07-Feb-07	3
ATGCCTGAGCAG	01_100	100	1	09-Feb-07	3
GTCGCTGTCTTC	01_101	101	1	12-Feb-07	3
AGTTCAGACGCT	01_112	112	1	19-Mar-07	4
ATCCTCAGTAGT	01_114	114	1	21-Mar-07	5
ATAGCTCCATAC	01_116	116	1	26-Mar-07	5
TAGTCGTCTAGT	01_117	117	1	28-Mar-07	5
ACATCACTTAGC	02_003	3	2	07-Jun-06	1
ACGGTGAGTGTC	02_020	20	2	12-Jul-06	1
AGGCTACACGAC	02_022	22	2	24-Jul-06	1
ACTCGATTCGAT	02_051	51	2	06-Oct-06	1
TAGCACACCTAT	02_061	61	2	30-Oct-06	1
CATATACTCGCA	02_073	73	2	27-Nov-06	1
AGTTCTACGTCA	02_084	84	2	22-Dec-06	1
ACTTGTAGCAGC	02_090	90	2	12-Jan-07	1
AGTGAGAGAAGC	02_095	95	2	24-Jan-07	2
ATAATCTCGTCG	02_098	98	2	31-Jan-07	2
ATATCGCTACTG	02_101	101	2	07-Feb-07	2
ATCGCGGACGAT	02_103	103	2	12-Feb-07	2
ATCTTAGACTGC	02_106a	106	2	16-Feb-07	3
GTCTTCGTCGCT	02_108	108	2	18-Feb-07	3

Table A7.1 Continued

Barcode	Sputum ID	Sample No	Patient	Date	Exacerbation
ATGCGTAGTGCG	02_109a	109	2	19-Feb-07	3
ATGTGTCGACTT	02_111	111	2	21-Feb-07	3
TAGCGACATCTG	02_112	112	2	22-Feb-07	3
AGTGCGATGCGT	02_115	115	2	25-Feb-07	3
ATACACGTGGCG	02_118	118	2	28-Feb-07	3
ATATGCCAGTGC	02_122	122	2	07-Mar-07	4
ATCGCTCGAGGA	02_125	125	2	14-Mar-07	4
ATGACTCATTCG	02_128	128	2	21-Mar-07	4
AGATGTTCTGCT	02_131	131	2	28-Mar-07	5
ACTAGCTCCATA	02_137	137	2	16-Apr-07	5
CATCATGAGGCT	02_153	153	2	25-May-07	5
TACCGCTAGTAG	03_017	17	3	17-Jul-06	1
ATTCTGTGAGCG	03_030	30	3	16-Aug-06	1
TAGACTGTACTC	03_040	40	3	18-Sep-06	1
ATCTACTACACG	03_054	54	3	20-Oct-06	1
ATGTCACCGTGA	03_063	63	3	17-Nov-06	1
TAGCGGATCACG	03_072	72	3	08-Dec-06	1
GTCTCTCTACGC	03_085	85	3	15-Jan-07	1
AGAGTAGCTAAG	03_110	110	3	19-Mar-07	1
GTTCGCGTATAG	03_135	135	3	25-May-07	1
GTTGACGACAGC	05_002	2	4	14-Jul-06	1
TACGGTATGTCT	05_012	12	4	07-Aug-06	1
AGCTATCCACGA	05_014	14	4	30-Aug-06	2
ACGATGCGACCA	05_016	16	4	04-Sep-06	2
AGATCTCTGCAT	05_018	18	4	06-Sep-06	2
ACTGTGACTTCA	05_021	21	4	13-Sep-06	3
ACTACGTGTGGT	05_023	23	4	15-Sep-06	3
ACGCGATACTGG	05_024	24	4	17-Sep-06	3
ACATGTCACGTG	05_026	26	4	19-Sep-06	3
ACACGAGCCACA	05_028	28	4	21-Sep-06	3
AGTCACATCACT	05_030	30	4	23-Sep-06	3
AGCGTAGGTCGT	05_034	34	4	27-Sep-06	3
AGATCGGCTCGA	05_036	36	4	29-Sep-06	3
ACTGTCGAAGCT	05_038	38	4	01-Oct-06	3
ACTACAGCCTAT	05_040	40	4	03-Oct-06	3
ACGCAACTGCTA	05_041	41	4	04-Oct-06	3
TACACACATGGC	05_044	44	4	09-Oct-06	3
ACACATGTCTAC	05_045	45	4	11-Oct-06	3
AGTAGTATCCTC	05_047	47	4	13-Oct-06	3
AGCGCTGATGTG	05_049	49	4	15-Oct-06	3
AGATACACGCGC	05_051	51	4	17-Oct-06	3

Table A7.1 Continued

Barcode	Sputum ID	Sample No	Patient	Date	Exacerbation
ACTGTACGCGTA	05_053	53	4	19-Oct-06	3
ACGTTAGCACAC	05_055	55	4	21-Oct-06	3
ACGATGCGACCA	05_060	60	4	27-Oct-06	4
ACATCACTTAGC	05_062	62	4	01-Nov-06	4
ACACACTATGGC	05_065	65	4	08-Nov-06	4
AGTACTGCAGGC	05_068	68	4	15-Nov-06	4
AATCGTGACTCG	05_079	79	4	01-Dec-06	5
ATACGTCTTCGA	05_088	88	4	08-Jan-07	5
AGAACACGTCTC	05_098	98	4	05-Feb-07	5
GTTAGAGCACTC	05_112	112	4	01-Mar-07	5
ACGTTAGCACAC	05_125	125	4	28-Mar-07	5
AGCACACCTACA	05_130	130	4	27-Apr-07	5
AACTGTGCGTAC	05_138	138	4	21-May-07	5
ATGCCTGAGCAG	05_139	139	4	23-May-07	5
TACATCACCACA	06_006	6	5	26-Jul-06	1
ACGCTATCTGGA	06_021	21	5	13-Sep-06	1
ACAGAGTCGGCT	06_030	30	5	13-Oct-06	1
ATCGTACAACTC	06_061	61	5	10-Jan-07	1
AGCATATGAGAG	06_067	67	5	29-Jan-07	1
AGTCTACTCTGA	06_072	72	5	12-Feb-07	1
AGTCCATAGCTG	06_079	79	5	28-Feb-07	2
ACACGGTGTCTA	06_081	81	5	05-Mar-07	2
ATCACTAGTCAC	06_082	82	5	09-Mar-07	2
ACATTCAGCGCA	06_083	83	5	14-Mar-07	2
ACGCGCAGATAC	06_086a	86	5	20-Mar-07	2
ACTTGTAGCAGC	06_087	87	5	21-Mar-07	3
AGATGTTCTGCT	06_089	89	5	23-Mar-07	3
AGCTCCATACAG	06_090	90	5	24-Mar-07	3
ACACTAGATCCG	06_094	94	5	28-Mar-07	3
ACCACATACATC	06_096	96	5	30-Mar-07	3
ACCACATACATC	06_100	100	5	13-Apr-07	4
ACGCTATCTGGA	06_102	102	5	18-Apr-07	4
ACTATTGTCACG	06_112	112	5	30-May-07	5
ACGAGTGCTATC	06_125	125	5	09-Jul-07	5
TAAGCGCAGCAC	07_015	15	6	06-Sep-06	1
TAGCGACATCTG	07_036	36	6	13-Oct-06	1
AGCGACTGTGCA	07_053	53	6	22-Nov-06	2
AGAGTAGCTAAG	07_056	56	6	29-Nov-06	2
ACTGACAGCCAT	07_059	59	6	06-Dec-06	2
ACGTGAGAGAAT	07_062	62	6	12-Dec-06	2
ACAGTGCTTCAT	07_065	65	6	15-Dec-06	3
CATGGCTACACA	07_066	66	6	17-Dec-06	3
AATCAGTCTCGT	07_067	67	6	18-Dec-06	3

Table A7.1 Continued

Barcode	Sputum ID	Sample No	Patient	Date	Exacerbation
AGGTGTGATCGC	07_068	68	6	19-Dec-06	3
AGCCATACTGAC	07_071	71	6	27-Dec-06	4
ACTCTTCTAGAG	07_073	73	6	05-Jan-07	4
AGAGCAAGAGCA	07_075	75	6	10-Jan-07	4
TACTAATCTGCG	07_077	77	6	15-Jan-07	4
ATCTGAGCTGGT	07_085	85	6	02-Feb-07	5
CATGAGTGCTAC	07_094	94	6	23-Feb-07	5
ATGCGTAGTGCG	07_122	122	6	20-Apr-07	5
TAACTCTGATGC	07_137	137	6	30-May-07	5
TAGTCGTCTAGT	08_022	22	7	18-Sep-06	1
TAACAGTCGCTG	08_033	33	7	18-Oct-06	1
TAGAGAGAGTGG	08_086	86	7	02-Mar-07	1
TAGATAGCAGGA	08_097	97	7	28-Mar-07	1
TACTTACTGCAG	08_109	109	7	30-Apr-07	1
GTGTCTACATTG	08_121	121	7	01-Jun-07	1
ACAGTTGCGCGA	09_005	5	8	31-Jul-06	1
ACTGATCCTAGT	09_009	9	8	09-Aug-06	1
ACACTAGATCCG	09_017	17	8	30-Aug-06	1
AGTACTGCAGGC	09_027	27	8	22-Sep-06	1
AAGAGATGTCGA	09_047	47	8	27-Oct-06	1
ATGGCAGCTCTA	09_055	55	8	15-Nov-06	2
ATTATCGTGCAC	09_061	61	8	29-Nov-06	2
TAGCGGATCACG	09_064	64	8	06-Dec-06	2
AGTGGATGCTCT	09_066	66	8	08-Dec-06	3
ATACAGAGCTCC	09_068	68	8	10-Dec-06	3
ATCACGTAGCGG	09_071	71	8	12-Dec-06	3
ATCGTACAACTC	09_075	75	8	17-Dec-06	3
ATGAGACTCCAC	09_076	76	8	18-Dec-06	4
ATGGCGTGCACA	09_078	78	8	22-Dec-06	4
ATTCTGTGAGCG	09_080	80	8	03-Jan-07	4
TAGCTCGTAACT	09_082	82	8	08-Jan-07	4
AGTGTCACGGTG	09_085	85	8	15-Jan-07	4
AGTAGTATCCTC	09_087	87	8	22-Jan-07	5
GTCTCATGTAGG	09_120	120	8	21-Mar-07	5
TAGTGTGCTTCA	09_133	133	8	08-Apr-07	5
ACCTCGATCAGA	09_156	156	8	17-May-07	5
AGTCTCGCATAT	09_160	160	8	31-May-07	5
ACGTACTCAGTG	09_163	163	8	06-Jun-07	5
AGCAGTCGCGAT	09_303	303	8	27-Jun-07	5
ATGTGTCGACTT	10_001	1	9	16-Aug-06	1
CATATCGCAGTT	10_063	63	9	05-Jan-07	1
ACAGCAGTGGTC	10_067	67	9	15-Jan-07	1
CATCGTATCAAC	10_083	83	9	21-Feb-07	1

Table A7.1 Continued

Barcode	Sputum ID	Sample No	Patient	Date	Exacerbation
TAGTGCTGCGTA	10_124	124	9	21-May-07	1
GTTGTATACTCG	10_130	130	9	06-Jun-07	1
ACCTCGATCAGA	10_138	138	9	25-Jun-07	2
ACGTACTCAGTG	10_141	141	9	02-Jul-07	2
ACTCGCACAGGA	10_144	144	9	09-Jul-07	2
AGAGAGCAAGTG	10_147	147	9	15-Jul-07	3
AGCATATGAGAG	10_149	149	9	17-Jul-07	3
AGGCTACACGAC	10_150	150	9	18-Jul-07	3
AAGCTGCAGTCG	10_153	153	9	23-Jul-07	3
ACAGCTAGCTTG	10_157	157	9	01-Aug-07	4
ACCTGTCTCTCT	10_160	160	9	08-Aug-07	4
ACGTCTGTAGCA	10_162	162	9	13-Aug-07	4
ATGAGACTCCAC	11_006	6	10	18-Sep-06	1
CAACTCATCGTA	11_025	25	10	01-Nov-06	2
ATACAGAGCTCC	11_027	27	10	06-Nov-06	2
ATGTCACCGTGA	11_028	28	10	08-Nov-06	2
ATGCAGCTCAGT	11_032	32	10	17-Nov-06	2
ATCTGAGCTGGT	11_035	35	10	24-Nov-06	2
ATCCGATCACAG	11_036	36	10	25-Nov-06	3
ATACTCACTCAG	11_040	40	10	29-Nov-06	3
AGTTAGTGCGTC	11_041	41	10	30-Nov-06	3
TAGGTATCTCAC	11_045	45	10	04-Dec-06	3
CAACTATCAGCT	11_046	46	10	05-Dec-06	3
ATGTACGGCGAC	11_051	51	10	13-Dec-06	4
ATGCACTGGCGA	11_055	55	10	22-Dec-06	4
ATCTCTGGCATA	11_057	57	10	03-Jan-07	4
TACAGTCTCATG	11_113	113	10	09-Apr-07	5
GTGATAGTGCCG	11_127	139	10	09-May-07	5
ATGATCGAGAGA	11_154	154	10	13-Jul-07	5
GTCTATCGGAGT	11_163	163	10	15-Aug-07	5
ATCTGGTGCTAT	12_002	2	11	08-Oct-06	1
AGTGGATGCTCT	12_023	23	11	14-Nov-06	1
ACTCAGATACTC	12_037	37	11	10-Jan-07	2
AGCGAGCTATCT	12_038	38	11	11-Jan-07	2
AGACGTGCACTG	12_040	40	11	17-Jan-07	2
AGCAGCACTTGT	12_042	42	11	21-Jan-07	2
AGCTTGACAGCT	12_046	46	11	31-Jan-07	2
AACTGTGCGTAC	12_048	48	11	02-Feb-07	3
ACAGAGTCGGCT	12_050	50	11	04-Feb-07	3
ACCGCAGAGTCA	12_051	51	11	05-Feb-07	3
ACGGTGAGTGTC	12_054	54	11	08-Feb-07	3
ACTCGATTCGAT	12_056	56	11	10-Feb-07	3
AGACTGCGTACT	12_058	58	11	12-Feb-07	3

Table A7.1 Continued

Barcode	Sputum ID	Sample No	Patient	Date	Exacerbation
AGCAGTCGCGAT	12_061	61	11	15-Feb-07	3
AGGACGCACTGT	12_064	64	11	22-Feb-07	4
AAGAGATGTCGA	12_066	66	11	28-Feb-07	4
ACTGTACGCGTA	12_067	67	11	01-Mar-07	4
ACAGCAGTGGTC	12_069	69	11	07-Mar-07	4
ATCACGTAGCGG	12_082	82	11	23-Mar-07	5
ACACACTATGGC	12_083	83	11	15-Apr-07	5
ACTCGCACAGGA	12_108	108	11	06-Jun-07	5
ACCGCAGAGTCA	12_110	110	11	27-Sep-07	5
AGTGTCACGGTG	12_115	115	11	10-Oct-07	5
ACGCGCAGATAC	13_002	2	12	06-Oct-06	1
AGTTCAGACGCT	13_011	11	12	27-Oct-06	1
TACGTGTACGTG	13_038	38	12	20-Dec-06	1
AGACTGCGTACT	13_051	51	12	24-Jan-07	1
GTGGCGATACAC	13_060	60	12	14-Feb-07	1
ATACGTCTTCGA	13_068	68	12	05-Mar-07	2
ATCACTAGTCAC	13_071	71	12	12-Mar-07	2
ATCCTCAGTAGT	13_072	72	12	14-Mar-07	2
ATCTACTACACG	13_074	74	12	19-Mar-07	2
ATGATCGAGAGA	13_077	77	12	26-Mar-07	2
ATGGTCTACTAC	13_079	79	12	28-Mar-07	3
CAACACGCACGA	13_081	81	12	30-Mar-07	3
TAGCTGAGTCCA	13_083	83	12	01-Apr-07	3
AGTGTTCGATCG	13_086	86	12	04-Apr-07	3
AGCTCTCAGAGG	13_090	90	12	18-Apr-07	4
ATACTATTGCGC	13_091	91	12	20-Apr-07	4
ATCAGGCGTGTG	13_094	94	12	27-Apr-07	4
ACGTGCCGTAGA	13_099	99	12	11-May-07	5
TACGATGACCAC	13_106	106	12	30-May-07	5
ATGGTCTACTAC	13_113	113	12	11-Jun-07	5
ACATTCAGCGCA	13_142	142	12	17-Aug-07	5
ACACGGTGTCTA	13_149	149	12	05-Sep-07	5

Table A7.2 Species-level identities of detected bacterial taxa identified from 237 sputum samples collected from CF patients.

Given the length of the ribosomal sequences analysed, these identities should be considered putative. Ae denotes aerobe and An, Anaerobe. Only strict anaerobes were classified as anaerobes, whereas aerobes, facultative anaerobes, and microaerophiles were classified as aerobes.

Class	Family	Taxon name	Common or Rare	Anaerobe / Aerobe
Actinobacteria	Actinomycetaceae	*Actinobaculum massiliense*	Rare	Ae
		Actinomyces graevenitzii	Rare	Ae
		Actinomyces naeslundii	Rare	Ae
		Actinomyces odontolyticus	Rare	Ae
	Bifidobacteriaceae	*Scardovia inopinata*	Rare	Ae
	Coriobacteriaceae	*Atopobium parvulum*	Rare	An
	Corynebacteriaceae	*Corynebacterium durum*	Rare	Ae
		Corynebacterium matruchotii	Rare	Ae
		Corynebacterium tuberculostearicum	Rare	Ae
	Microbacteriaceae	*Microbacterium paraoxydans*	Rare	Ae
	Micrococcaceae	*Arthrobacter tumbae*	Rare	An
		Rothia mucilaginosa	Rare	Ae
	Nocardiaceae	*Nocardia cyriacigeorgica*	Rare	An
	Propionibacteriaceae	*Propionibacterium acidifaciens*	Rare	An
		Propionibacterium acnes	Rare	An
		Propionibacterium propionicum	Rare	Ae
Alphaproteobacteria	Brucellaceae	*Ochrobactrum anthropi*	Rare	Ae
	Caulobacteraceae	*Brevundimonas vesicularis*	Rare	Ae
	Rhodobacteraceae	*Paracoccus yeei*	Rare	Ae
	Sphingomonadaceae	*Sphingobium amiense*	Rare	An
		Sphingomonas paucimobilis	Rare	An
Bacilli	Aerococcaceae	*Abiotrophia defectiva*	Rare	Ae
	Bacillaceae	*Bacillus subtilis*	Rare	Ae
		Gemella sanguinis	Rare	Ae
	Carnobacteriaceae	*Granulicatella adiacens*	Rare	Ae
	Lactobacillaceae	*Lactobacillus casei*	Rare	Ae
		Lactobacillus delbrueckii	Rare	Ae
		Lactobacillus frumenti	Rare	Ae
		Lactobacillus johnsonii	Rare	Ae

Table A7.2 Continued

Class	Family	Taxon name	Common or Rare	Anaerobe / Aerobe
		Lactobacillus salivarius	Rare	Ae
	Staphylococcaceae	*Staphylococcus aureus*	Rare	Ae
	Streptococcaceae	*Streptococcus agalactiae*	Rare	An
		Streptococcus mutans	Rare	An
		Streptococcus pneumoniae	Common	An
		Streptococcus sanguinis group	Common	An
Bacteroidia	Bacteroidaceae	*Bacteroides acidofaciens*	Rare	An
		Bacteroides cellulosilyticus	Rare	An
		Bacteroides intestinalis	Rare	An
		Bacteroides oleiciplenus	Rare	An
		Bacteroides uniformis	Rare	An
	Porphyromonadaceae	*Barnesiella intestinihominis*	Rare	An
		Odoribacter laneus	Rare	an
		Paludibacter propionicigenes	Rare	An
		Parabacteroides distasonis	Rare	Ae
		Parabacteroides goldsteinii	Rare	Ae
		Porphyromonas catoniae	Rare	An
		Porphyromonas endodontalis	Rare	An
		Porphyromonas somerae	Rare	An
		Tannerella forsythia	Rare	An
	Prevotellaceae	*Prevotella bivia*	Rare	An
		Prevotella buccae	Rare	An
		Prevotella copri	Rare	An
		Prevotella denticola	Rare	An
		Prevotella enoeca	Rare	An
		Prevotella histicola	Rare	An
		Prevotella loescheii	Rare	An
		Prevotella maculosa	Rare	An
		Prevotella melaninogenica	Common	An
		Prevotella nanceiensis	Rare	An
		Prevotella nigrescens	Rare	An
		Prevotella oralis	Rare	An
		Prevotella oris	Rare	An
		Prevotella oulorum	Rare	An
		Prevotella pallens	Rare	An
		Prevotella paludivivens	Rare	An
		Prevotella shahii	Rare	An
		Prevotella tannerae	Rare	An
	Rikenellaceae	*Alistipes finegoldii*	Rare	An

Table A7.2 Continued

Class	Family	Taxon name	Common or Rare	Anaerobe / Aerobe
Betaproteobacteria	Alcaligenaceae	*Achromobacter xylosoxidans*	Rare	Ae
		Advenella mimigardefordensis	Rare	Ae
	Burkholderiaceae	*Burkholderia cepacia complex*	Rare	Ae
		Lautropia mirabilis	Rare	Ae
		Ralstonia mannitolilytica	Rare	Ae
	Burkholderiales *incertae sedis*	*Aquabacterium fontiphilum*	Rare	Ae
	Comamonadaceae	*Comamonas testosteroni*	Rare	Ae
		Curvibacter lanceolatus	Rare	Ae
		Delftia acidovorans	Rare	Ae
	Neisseriaceae	*Kingella oralis*	Rare	Ae
		Neisseria mucosa	Rare	Ae
		Neisseria oralis	Rare	Ae
	Sutterellaceae	*Parasutterella excrementihominis*	Rare	An
Clostridia	Clostridiaceae	*Anaerococcus octavius*	Rare	An
		Clostridium aerotolerans	Rare	An
		Clostridium aldenense	Rare	An
		Clostridium algidixylanolyticum	Rare	An
		Clostridium bolteae	Rare	An
		Clostridium celerecrescens	Rare	An
		Clostridium cellobioparum	Rare	An
		Clostridium clostridioforme	Rare	An
		Clostridium hathewayi	Rare	An
		Clostridium indolis	Rare	An
		Clostridium lavalense	Rare	An
		Clostridium orbiscindens	Rare	An
		Clostridium papyrosolvens	Rare	
		Clostridium piliforme	Rare	An
		Clostridium populeti	Rare	An
		Clostridium proteoclasticum	Rare	An
		Clostridium scindens	Rare	An
		Clostridium subterminale	Rare	An
		Clostridium sufflavum	Rare	An
		Clostridium tertium	Rare	An
		Finegoldia magna	Rare	An
		Mogibacterium neglectum	Rare	An

Table A7.2 Continued

Class	Family	Taxon name	Common or Rare	Anaerobe / Aerobe
		Parvimonas micros	Rare	An
		Peptoniphilus lacrimalis	Rare	An
	Eubacteriaceae	*Eubacterium brachy*	Rare	An
		Eubacterium sulci	Rare	An
		Eubacterium ventriosum	Rare	An
	Lachnospiraceae	*Anaerostipes butyraticus*	Rare	An
		Blautia obeum	Rare	An
		Catonella morbi	Rare	An
		Howardella ureilytica	Rare	An
		Lachnoanaerobaculum orale	Rare	An
		Oribacterium sinus	Rare	An
		Robinsoniella peoriensis	Rare	An
		Shuttleworthia satelles	Rare	An
		Stomatobaculum longum	Rare	An
	Peptococcaceae	*Peptococcus niger*	Rare	An
	Peptostreptococcaceae	*Peptostreptococcus stomatis*	Rare	An
	Ruminococcaceae	*Pseudoflavonifractor capillosus*	Rare	Ae
		Ruminococcus flavefaciens	Rare	An
Deltaproteobacteria	Bdellovibrionaceae	*Vampirovibrio chlorellavorus*	Rare	Ae
Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*	Rare	Ae
		Campylobacter curvus	Rare	Ae
		Campylobacter showae	Rare	Ae
Erysipelotrichia	Erysipelotrichaceae	*Eubacterium cylindroides*	Rare	An
Flavobacteria	Flavobacteriaceae	*Capnocytophaga granulosa*	Rare	Ae
		Capnocytophaga ochracea	Rare	Ae
		Capnocytophaga sputigena	Rare	Ae
		Chryseobacterium indologenes	Rare	Ae
Fusobacteria	Fusobacteriaceae	*Fusobacterium necrophorum*	Rare	An
		Fusobacterium nucleatum	Rare	An
	Leptotrichiaceae	*Leptotrichia buccalis*	Rare	Ae
		Sneathia sanguinegens	Rare	Ae
Gammaproteobacteria	Cardiobacteriaceae	*Cardiobacterium valvarum*	Rare	Ae
	Enterobacteriaceae	*Enterobacter cowanii*	Rare	Ae
		Proteus mirabilis	Rare	Ae
		Yersinia frederiksenii	Rare	Ae
	Moraxellaceae	*Acinetobacter johnsonii*	Rare	Ae
		Acinetobacter lwoffii	Rare	Ae
		Moraxella nonliquefaciens	Rare	Ae

Table A7.2 Continued

Class	Family	Taxon name	Common or Rare	Anaerobe / Aerobe
		Moraxella osloensis	Rare	Ae
	Pasteurellaceae	*Haemophilus influenzae*	Rare	Ae
		Haemophilus parainfluenzae	Rare	Ae
	Pseudomonadaceae	*Pseudomonas aeruginosa*	Common	Ae
		Pseudomonas fragi	Rare	Ae
		Pseudomonas pseudoalcaligenes	Rare	Ae
	Xanthomonadaceae	*Stenotrophomonas maltophilia*	Rare	An
Mollicutes	Mycoplasmataceae	*Mycoplasma hominis*	Rare	Ae
		Mycoplasma salivarium	Rare	An
Negativicutes	Veillonellaceae	*Anaeroglobus geminatus*	Rare	An
		Dialister invisus	Rare	Ae
		Dialister micraerophilus	Rare	Ae
		Dialister pneumosintes	Rare	Ae
		Megasphaera micronuciformis	Rare	An
		Schwartzia succinivorans	Rare	An
		Selenomonas artemidis	Rare	An
		Selenomonas noxia	Rare	An
		Veillonella parvula	Common	An
		Veillonella ratti	Rare	An
Sphingobacteria	Chitinophagaceae	*Sediminibacterium salmoneum*	Rare	Ae
	Sphingobacteriaceae	*Sphingobacterium spiritivorum*	Rare	Ae
Spirochaetes	Spirochaetaceae	*Treponema denticola*	Rare	Ae

Chapter 8: Discussion

General Discussion

Chronic respiratory infections are the main cause of morbidity and mortality in individuals suffering from cystic fibrosis. Improvements in the treatments available for CF has increased the median predicted survival age of these patients to 41.1 years of age (1). However, in order to increase this further a much greater understanding of the microbial community associated with this disease is required. The use of culture independent technologies in recent years has revealed a complex and diverse bacterial community within the CF lung, yet it is still unclear how this community is related to clinical outcomes.

The main aim of this dissertation was to investigate the relationship between the bacterial community within the CF lung and host related clinical factors. By examining these relationships it was anticipated that this work would uncover potentially important factors that may influence treatment regimes and ultimately improve prognosis. This was achieved through high throughput sequencing of the bacterial community extracted from CF sputum. In this chapter the findings generated throughout this project will be discussed in relation to their potential impact on CF research.

8.1 Reducing bias

The aim of every researcher is to gain an unbiased representation of their chosen study system. It is possible to introduce bias at every level of sampling; from sample collection and storage, to DNA extraction, PCR and sequencing. As such, in order to be confident in the observed results efforts must be made to both understand and reduce the effect of potential bias. Consequently, throughout this study standard procedures including; extraction controls, both positive and negative PCR controls and reagent controls were included in every study and sequencing run (2). While these measures controlled for contamination during sequencing these actions alone would not account for all areas of potential bias; the most prevalent perhaps being the presences of extracellular DNA and sample handling.

8.1.1 Extracellular DNA

The presence of extracellular DNA within the CF lung is inevitable due to immune responses and the continual administration of antibiotics, as well as the lack of ability to clear mucus from the airways. As a result, PMA was identified by Rogers *et al* (2008) as a potentially important method of reducing the impact of this extracellular DNA when investigating the bacterial community within the CF lung (3). Chapter 3 supports the findings of Roger *et al* (2013), confirming that failure to account for the presence of extracellular DNA or DNA from dead or damaged cells can result in a misrepresentation of the bacterial community either by an under or over estimation of the bacterial diversity (4).

The use of PMA was also found to have significantly affected the abundance of common organisms, in particular the recognised CF pathogen *Pseudomonas aeruginosa*. These results show that failure to treat samples with PMA prior to culture independent analysis may result in the masking of important changes within the community. This in turn could lead to misrepresentative data and consequentially effect treatment decisions. As a result of these findings, PMA treatment was used throughout this project.

In Chapter 7, the use of PMA was particularly important as it allowed an unbiased overview of the effect of perturbations on the bacterial community within the lung. Without the use of PMA the impact of antibiotic treatment may have been masked by the amplification of non-viable DNA from bacteria killed as a result of aggressive intervention. Consequently, studies such as that carried out by Fodor et al (2012), which failed to include PMA treatment, may have missed important changes in the overall community structure with antibiotic intervention for CFPE (5).

8.1.2 Sample handling

From the investigation of CF lung infections by culture based microbiology of sputum, it has been recognised that the storage and handling of clinical samples is vital to obtaining a representative depiction of the bacterial community (6). While studies have been undertaken to determine the optimal storage and transport of sputum for culture, few have investigated

the impact of sample handling prior to the use of culture independent techniques. As moves are made towards the introduction of these techniques for clinical diagnostics it is becoming more important than ever that samples are handled appropriately allowing the most accurate representation of the true bacterial community to be achieved (7).

Due to the practicalities involved with processing samples for culture independent analysis samples are collected and stored at -80°C until processing. This ensures rapid stabilisation of the bacterial community from the time of freezing and for as long as necessary. In Chapter 4, the bacterial community within sputum samples, frozen at intervals over a 72 hour period was assessed revealing significant changes the abundance of anaerobic taxa after storage at room temperature for more than 12 hours. Therefore, recommendations were made that respiratory samples should be stabilised to -80°C within 12 hours of collection in order to avoid significant changes in bacterial community composition.

Obtaining large sample sets of respiratory samples with associated metadata is a huge undertaking both for researchers and clinicians. Consequently, the use of previously collected samples for multiple studies can provide a more practical option. As such, many clinics are collating large detailed biobanks of samples for research purposes. While these biobanks represent a wealth of potential information for researchers investigating a wide range of factors relating to CF lung infections they represent another avenue of bias.

Once stored at -80°C, common practice dictated that sputum may be defrosted and subsampled once, after which they should no longer be refrozen and sub-sampled at a later date without occurring significant changes to the microbial community. However, no studies actually examined this relationship. By carrying out 6 freeze thaw cycles on sputum samples from eight patients, changes in the bacterial community as a result of these cycles were assessed in Chapter 5. No change in the whole bacterial community similarity was observed over 6 freeze-thaw cycles however, after 4 cycles the rare community fell below the level expected for within sample variation. As such sub-sampling should be carried out a maximum of 3 times in order to obtain an accurate representation of the bacterial community

within the respiratory tract. In the case of large volumes of sputum a more practical solution would be to aliquot the sample avoiding unnecessary freeze-thaw cycles.

Chapters 3, 4 and 5, by assessing areas of potential bias, allowed the identification of guidelines for the storage and treatment of sputum samples prior to DNA based analysis.

These guidelines directed clinicians and researchers to store samples as soon as possible after collection (within a 12 hour window) to only allow minimal freeze-thaw cycles when sub-sampling sputum and to include PMA treatment in the processing of all sputum samples. These clearly researched guidelines provided strong evidence that the results presented in this thesis were a true, unbiased representation of the bacterial community.

8.2 Understanding the metacommunity through data partitioning

The idea of the metacommunity was first introduced into the analysis of bacterial communities within the CF lung by van der Gast *et al* (2011). By partitioning the bacterial community into the most common and abundant taxa, and those that are rare and transient, it became possible to uncover how the community was effected by a variety of factors. Throughout this work partitioning of the bacterial community has been used to identify patterns and determine the taxa responsible for the changes observed. Across all studies the use of partitioning revealed potentially important factors relating to the bacterial community that would have been missed had this partitioning not been implemented, this was particularly true of the rare OTUs.

When considering the reduction of bias in the initial chapters discussed previously, failure to partition the bacterial community would have resulted in less conserved estimates of time to freezing and a failure to identify any change in the community over the 6 freeze-thaw cycles. This could have important consequences for future studies, where rare bacterial taxa may have been misrepresented within the community.

The importance of partitioning data when analysing the bacterial community in relation to clinical factors was identified in the results of Chapters 6 and 7. For example in Chapter 6,

had partitioning not been undertaken the relationship between the bacterial community and gender would not have been identified. While this relationship requires further investigation, the identification of a bacterial community relationship with gender could be of enormous value in the treatment of CF lung disease. Further, when considering the bacterial community over the course of an exacerbation, partitioning of OTUs provided valuable insight into the dynamics of the bacterial community. It was shown that rate of species turnover was driven by the rare taxa, while patterns in community composition were controlled by the common organisms. Partitioning the data in this study allowed patterns of resilience in the rare to be revealed, a relationship unobserved by other studies.

8.3 Relating bacterial community to clinical factors on a large scale

Several relationships have been observed relating interpatient differences in the bacterial community to clinical factors however, the studies presenting these findings have lacked statistical power due the small sample sizes involved. For example both Delhaes *et al* (2012) and van der Gast *et al* (2011) identified a significant relationship between bacterial diversity and lung function however this relationship was based on the bacterial community of 4 and 14 patients respectively (8, 9). Using bacterial community data from 292 individual CF patients previously described relationships, such as that between FEV_1 (Forced expiratory volume in 1 second) and bacterial diversity, were examined to determine if these relationships were genuine or simply an artefact of the limited sample size used in these studies.

Due to the huge size and complexity of the dataset investigated in Chapter 6, this thesis aimed to highlight preliminary findings. Consequently, throughout this section further work that could be undertaken to explore the relationships exposed by this study further are discussed.

8.3.1 Bacterial diversity and lung function

As explained to above, a significant relationship between lung function and the bacterial community has previously been revealed (8, 9). Through the investigation of this relationship using a much larger dataset it was revealed that while a significant relationship is present between bacterial diversity and FEV_1, the interpatient variation observed was such that the variance explained by this relationship was low. A significant correlation between bacterial community composition and lung function was also observed using Mantel tests, while partial Mantel tests revealed correlation between lung function and community composition when controlling for clinical status (stable or experiencing a CFPE), BMI and liver disease. A significant correlation between the rare taxa composition and FEV_1 was also found when controlling for clinic location.

Understanding the relationship between the bacterial community and lung function is important as FEV_1 is considered to be the best predictor for mortality in CF individuals (10). As a decrease in lung function is associated with poor clinical outcomes, it is important to understand how this prognosis is related to the bacterial community and other clinical factors. By understanding this relationship it may be possible to identify risk factors to reduced lung function and treat patients accordingly. Further work on this area would involve the use of detailed multivariate analysis, where FEV_1 and the 4 related clinical factors, identified using partial mantel testing, are used as explanatory variables. Additional investigation into the effect of individual bacterial taxa on the clinical factors may also reveal important relationships relating to disease outcomes.

In addition to the relationships identified above, variation in the relationship between diversity and FEV_1, could be further explained by investigating the relationship between the bacterial diversity, lung function and treatment regimes. This analysis could not only include investigation into the effect of antibiotic treatment regimes but also the effect of different methods for the administration as well as other treatments including, steroids and mucolytics.

8.3.2 CFTR genotype and the bacterial community

CFTR (CF transmembrane regulator) genotype has been suggested as a potentially important factor in disease progression with some mutations found to correlate with more severe disease than others (11). In accordance with the knowledge of mutation frequency, at least one gene showed the ΔF508 mutation in the majority of patients, alongside this a large variation in other mutations were observed. As a result 3 groups were outlined; ΔF508 homozygotes, ΔF508 heterozygotes and other mutations. However, only ANOSIM revealed a significant difference in community composition between patients with homozygous and heterozygous ΔF508 mutations (P=0.03).

While this is not compelling evidence for genotype based differences in bacterial community, it does not rule out the relationship. By partitioning the ΔF508 heterozygotes into those associated with the severe CF phenotype and those associated with a milder phenotype, potential differences in the bacterial community could be uncovered.

8.3.3 The gender gap and the bacterial community

A so called gender gap has historically been observed been CF individuals, with females showing significantly worse prognosis than males (12, 13). Although this gap has become less pronounced more recently (14), an investigation was carried out to understand if the bacterial community was a factor in the disparity between genders. This analysis revealed that female patients had significantly lower bacterial diversity; while ANOSIM revealed a significant difference in the community composition of the rare taxa. Further analysis of the organisms responsible for the disparity between females and males revealed that the recognised CF pathogens *Pseudomonas aeruginosa* and *Staphylococcus aureus* had a higher mean abundance in female patients.

The relationship between the bacterial community and gender warrants further investigation to explore the potential reasons for the differences in community diversity and composition observed. Firstly, the relationship between *P. aeruginosa* and *S. aureus* mean abundance and gender requires further study. It has been previously shown that co-infection with these organisms is associated with poor prognosis (15, 16), therefore examination of this relationship could potentially reveal an important reason for the gender gap. In addition, a

more in depth study of how the rare communities differ between genders may provide other potentially important species which contribute to the prognosis of females.

8.3.4 The bacterial community by CF centre

The relationship between the bacterial community and CF centres is complex and multifaceted involving geographic location, treatment regimes employed and environment surrounding the centre catchment area (e.g. urban or rural). As a result differences between CF centres are difficult to elucidate. In an attempt to identify any differences in the bacterial community between centres, bacterial diversity and community composition were explored. Although differences in bacterial diversity were observed, post-hoc testing revealed these differences were not consistent or ubiquitous between centres, rather they were more likely to be an artefact of interpatient variation than true differences between centres.

Mantel test results suggested that patients from CF centres closer in geographic distance would show more similar bacterial communities. However, ANOSIM showed little evidence that this was the case. Even so it did reveal that the common taxa were more conserved between sites than those considered to be rare. These data suggested therefore that it is the rare community that drives differences by centre or geographical location. Further study would be required to determine if these transient organisms are associated with the surrounding environment of the centre.

8.4 Bacterial community dynamics as a result of pulmonary exacerbation

CF pulmonary exacerbations (CFPE) are associated with more rapid disease progression, aggressive antibiotic treatment and hospital stays, making them both medically and economically important phenomena (17). Using longitudinal sampling the bacterial community within the CF lung was examined over the full course of a CFPE; from baseline,

leading into exacerbation, treatment, leading out and back to baseline. This full cycle approach allowed both the lead in and the recovery to be examined for the first time.

This study observed that the changes in the bacterial community were associated with changes in rare taxa. This was highlighted when investigating rates of species turnover as well as community composition over the study period. The rare group was shown to significantly change leading into, throughout and leading out of treatment for CFPE. It was also found to return to baseline levels within the 30 days after treatment had ceased, indicating resilience within the rare group.

Taxa partitioned in the common group were found to be resistant to changes due to CFPE treatment however, when the relative abundance of these OTUs were examined individually key taxa, *Prevotella melaninogenica* and *Veillonella parvula*, were identified as potential biomarkers for CFPE. *V. parvula* was considered to be the most appropriate organism for further investigation due to the increase in abundance of this OTU prior to the initiation of treatment for CFPE.

The association between CFPE and disease progression is a major concern in CF research, therefore the identification of CFPE is hugely important. With no consensus as to the clinical symptoms which indicate the start of CFPE, clinicians are required to make an antibiotic intervention judgement call based on worsening health. It is known that rapid intervention with antibiotic treatment results in the return of lung function to baseline states (18), however without a clear understanding of what causes a CFPE the initiation of rapid intervention is challenging. The identification of biomarkers for disease state could be the answer to this issue. By introducing a rapidly testable parameter for clinicians to identify a CFPE, a reduction in the use of unnecessary intervention could result. This in itself could provide one of the most important recent developments in CF therapy.

The potential importance of these biomarkers (in particular *V. parvula* due to its rapid increase in abundance prior to the antibiotic treatment) cannot be underestimated. To investigate the potential of *V. parvula,* further work is required over the course of CFPE, this

includes using qPCR to track this organism, allowing a truly quantitative indication of the bacterial abundance, as well as large scale screening.

8.5 Conclusions

This thesis examined the bacterial community ecology within the CF lung through the use of next generation sequencing technologies (Roche 454 and Illumina MiSeq). This work was carried out in two sections, the first examined the importance of reducing bias in order to obtain the most representative depiction of the bacterial community. The results from the initial work were used to underpin work carried out in the second half of this project, which focused on the relationship between the bacterial community and clinical factors.

Conclusions that can be drawn from this dissertation include;

- The use of PMA is required to obtain a true picture of the bacterial community within the CF lung.
- Sputum samples should be stored at -80°C within 12 hours of collection to avoid significant changes in the bacterial community.
- Sputum samples for culture independent analysis should go through no more than 3 freeze-thaw cycles to avoid significant changes in the rare taxa.
- FEV_1 is not a reliable indicator of bacterial diversity due to the levels of interpatient variation seen in CF patients.
- Significant differences are observed in the rare bacterial community between females and males.
- The common bacterial taxa are conserved between CF centres, however differences in the rare taxa are correlated with geographical distance.
- Over the course of a pulmonary exacerbation the common bacterial taxa show resistance while the rare taxa show resilience.
- *Veillonella parvula* has the potential to be a bioindicator of CFPE.

Through validated sampling methods this study has revealed several factors which have potential to inform treatment and improve prognosis. However, further research is required to investigate the full extent of this potential.

8.6 Future work

This project aimed to provide a greater understanding of the bacterial community within the CF lung. The complexity of the bacterial community throughout this investigation has informed the conclusion that while the bacterial community plays an important role in disease progression, assessment of the wider microbial diversity is required in order to understand CF lung disease progression more fully. Several studies have been published examining both bacteria and fungi within the CF lung (8, 19), however, no study to date has encompassed the whole microbiota including the bacterial, fungal and viral communities.

Although a wider more microbial based assessment of the CF lung is important, there is also a real requirement for a more multi-disciplinary approach to assessing how the microbial community relates to disease state. By combining microbial data with immunological data as well as clinical metadata, real insight into the complexities of infection can be uncovered. Further, the use of metatranscriptomics to investigate gene function could potentially uncover important insights into microbial activates as well as diversity in response to perturbations within the lung.

In order to gain meaningful data from studies of this nature, there is a requirement for large sample sets, to account for the degree of individual patient variation between individual patients. There is also a pressing need for the data sets to not just include adult patients but also young children (<10 years of age). Further, there is a need for large, longitudinal studies which sample patients over long periods, throughout times of stability and over the course of CFPE.

Despite the pressing need for studies of this nature and size to be carried out, the organisation and time involved in collecting samples for these sorts of investigations would

be a huge commitment for any research group, most likely requiring several collaborating research centres and CF clinics in order to obtain the sample sizes required. Despite the work involved, CF research groups have to move towards studies of this nature in order to fully understand CF disease progression and ultimately it treatment.

It must not be overlooked that CF makes up a very small percentage of the lower respiratory tract infections (LRTI) worldwide. LRTI and chronic obstructive pulmonary disease (COPD) are currently estimated by the world health organisation (WHO) to result in around 6.2 million deaths worldwide (3.1 million LRTI, 3.1 million COPD) (20). These numbers are staggering and underline the requirement for this type of research to be applied to a much wider range of pulmonary diseases, for example; COPD, bronchiectasis, bronchitis and severe asthma. Techniques and approaches presented throughout this thesis are already being applied to the bacterial communities of individuals with bronchiectasis and bronchitis (21, 22). The application of the knowledge gained in the study of CF to other respiratory disease has enormous potential to increase our understanding of lung disease and as a result reduce the associated mortality.

8.7 References

1. **Registry CFFP.** 2012. 2012 Annual Data Report. ©2014 Cystic Fibrosis Foundation.

2. **Salter S, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman N, Walker AW.** 2014. Reagent contamination can critically impact sequence-based microbiome analyses. bioRxiv **doi:** http://dx.doi.org/10.1101/007187

3. **Rogers GB, Stressmann FA, Koller G, Daniels T, Carroll MP, Bruce KD.** 2008. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diag Microbiol Infect Dis **62:**133-141.

4. **Rogers GB, Cuthbertson L, Hoffman LR, Wing PA, Pope C, Hooftman DA, Lilley AK, Oliver A, Carroll MP, Bruce KD, van der Gast CJ.** 2013. Reducing bias in bacterial community analysis of lower respiratory infections. ISME J **7:**697-706.

5. **Fodor A, Klem E, Gilpin D, Elborn J, Boucher R, Tunney M, Wolfgang M.** 2012. The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. PLOS One **7:**e45001.

6. **Health Protection Agency.** 2012. Investigation of Bronchoalveolar Lavage, Sputum and Associated Specimens. UK Standards for Microbiology Investigations**:**http://www.hpa.org.uk/SMI/pdf.

7. **Price K, Hampton T, Gifford A, Dolben E, Hogan D, Morrison H, Sogin M, O'Toole G.** 2013. Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. Microbiome **1:**27-37.

8. **Delhaes L, Monchy S, Frealle E, Hubans C, Salleron J, Leroy S, Prevotat A, Wallet F, Wallaert B, Dei-Cas E, Sime-Ngando T, Chabe M, Viscogliosi E.** 2012. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community--implications for therapeutic management. PLOS One **7:**e36313.

9. **van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD.** 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. ISME J **5:**780-791.

10. **Kerem E, Reisman J, Corey M, Canny GJ, Levison H.** 1992. Prediction of Mortality in Patients with Cystic Fibrosis. New Engl J Med **326:**1187-1191.

11. **Zielenski J, Tsui LC.** 1995. Cystic fibrosis: Genotypic and phenotypic variations. Annu Rev Genet **29:**777-807.

12. **Rosenfeld M, Davis R, FitzSimmons S, Pepe M, Ramsey B.** 1997. Gender gap in cystic fibrosis mortality. Am J Epidemiol **145:**794-803.

13. **McKone EF, Goss CH, Aitken ML.** 2006. CFtr genotype as a predictor of prognosis in cystic fibrosis*. Chest **130:**1441-1447.

14. **Verma N, Bush A, Buchdahl R.** 2005. Is there still a gender gap in cystic fibrosis? Chest **128:**2824-2834.

15. **Korgaonkar A, Trivedi U, Rumbaugh KP, Whiteley M.** 2013. Community surveillance enhances Pseudomonas aeruginosa virulence during polymicrobial infection. PNAS **110:**1059-1064.

16. **Sibley CD, Duan K, Fischer C, Parkins MD, Storey DG, Rabin HR, Surette MG.** 2008. Discerning the Complexity of Community Interactions Using a *Drosophila* Model of Polymicrobial Infections. PLOS Pathog **4:**e1000184.

17. **VanDevanter DR, O'Riordan MA, Blumer JL, Konstan MW.** 2010. Assessing time to pulmonary function benefit following antibiotic treatment of acute cystic fibrosis exacerbations. Resp Res **11**:137-144

18. **Sanders DB, Hoffman LR, Emerson J, Gibson RL, Rosenfeld M, Redding GJ, Goss CH.** 2010. Return of FEV(1) After Pulmonary Exacerbation in Children With Cystic Fibrosis. Pediatr Pulm **45:**127-134.

19. **Charlson ES, Diamond JM, Bittinger K, Fitzgerald AS, Yadav A, Haas AR, Bushman FD, Collman RG.** 2012. Lung-enriched Organisms and Aberrant Bacterial and Fungal Respiratory Microbiota after Lung Transplant. Am J Resp Crit Care **186:**536-545.

20. **Organisation WH.** 2008. The top 10 causes of death. www.who.int **Fact sheet number 310**.

21. **Rogers GB, van der Gast CJ, Cuthbertson L, Thomson SK, Bruce KD, Martin ML, Serisier DJ.** 2013. Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. Thorax **68:**731-737.

22. **van der Gast CJ, Cuthbertson L, Rogers GB, Pope C, Marsh RL, Redding GJ, Bruce KD, Chang AB, Hoffman LR.** 2014. Three clinically distinct chronic pediatric airway infections share a common core microbiota. Ann Am Thorac Soc **11:**1039-1048.